A NEW LOOK AT JOINT ATTENTION AND COMMON KNOWLEDGE

Barbora Siposova[1,2] and Malinda Carpenter[1,2]

[1]School of Psychology and Neuroscience, University of St Andrews, St Andrews, Scotland, UK

[2]Department of Developmental and Comparative Psychology, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

In press, *Cognition*

Correspondence should be addressed to Barbora Siposova, Department of Psychology, University of Warwick, Coventry CV4 7AL, UK, Email: siposova@protonmail.com

Abstract

Everyone agrees that joint attention is a key feature of human social cognition. Yet, despite over 40 years of work and hundreds of publications on this topic, there is still surprisingly little agreement on what exactly joint attention is, and how the jointness in it is achieved. Part of the problem, we propose, is that joint attention is not a single process, but rather it includes a cluster of different cognitive skills and processes, and different researchers focus on different aspects of it. A similar problem applies to common knowledge. Here we present a new approach: We outline a typology of social attention levels which are currently all referred to in the literature as joint attention (from monitoring to common, mutual, and shared attention), along with corresponding levels of common knowledge. We consider cognitive, behavioral, and phenomenological aspects of the different levels as well as their different functions, and a key distinction we make in all of this is second-personal vs. third-personal relations. While we focus mainly on joint attention and common knowledge, we also briefly discuss how these levels might apply to other 'joint' mental states such as joint goals.

Keywords: joint attention, common knowledge, second-person relations, communication, joint action, commitments

A New Look at Joint Attention and Common Knowledge

1. Introduction

From infancy through adulthood, joint attention is one of the most important skills in human social cognition. In joint attention we focus on things together with others (Mundy, Sullivan, & Mastergeorge, 2009). Thus, joint attention allows us to share experiences about the world with others, to coordinate our thoughts and behaviors, and to cooperate successfully with others (Tomasello, 1999; Trevarthen, 1979). It provides a foundation for early language development and social competence and facilitates social learning (Mundy, 2003). Individuals who lack the skills to follow and share attention with others have significant difficulties in relating to other people and sustaining relationships (Mundy & Newell, 2007).

Despite universal agreement about the importance of joint attention, however, there is still surprisingly little agreement on exactly what joint attention is and how it is achieved. Even now, more than 40 years after the first articles on joint attention were published (Bruner, 1974; Lewis, 1969; Scaife & Bruner, 1975), there is much variability in how different researchers define and use the term joint attention. For example, some focus mainly on behaviors and define joint attention simply as "looking where someone else is looking" (Butterworth, 1995, p. 29). Others go further and define it as "the intentional co-orientation of two or more organisms to the same locus" (Leavens & Racine, 2009, p. 241). Others bring in the idea of interest, defining joint attention as "the ability to coordinate attention toward a social partner and an object of mutual interest" (Bakeman & Adamson, 1984, p. 1278; see also Baron-Cohen, 1989; Camaioni, 1993; Mundy & Newell, 2007). Still others go further, considering what kind of awareness the individuals must have and arguing that to be in joint attention, the individuals must *know*

together that they are attending to the same thing: They must have common knowledge of this (e.g., Carpenter & Liebal, 2011; R. P. Hobson, 2005; Tomasello, 1995). On the surface, these definitions may all sound similar – they all involve two individuals attending to the same thing – but actually, as we will see, there are important differences among them; differences which may account for many misunderstandings and unnecessary debates in the literature (Carpenter & Call, 2013).

A related long-standing debate concerns how the *jointness* and common knowledge in joint attention (and joint action) are achieved. Common knowledge was initially formally characterized as an infinite iteration of knowledge states such as "I know that you know, you know that I know that you know, I know that you know that I know that you know", etc., ad infinitum (e.g., Lewis, 1969; Schiffer, 1972). There are definitional and terminological issues here as well: Some authors use the term "common knowledge" (Lewis, 1969) and others use the term "mutual knowledge" (Pinker, 2007; Schiffer, 1972) to describe this phenomenon, whereas others reserve "mutual knowledge" for the state in which you know X and I know X, but nothing else (Vanderschraaf & Sillari, 2014). Still others use "common knowledge" and "mutual knowledge" interchangeably (e.g., Clark, 1992, 1996). Regardless of which term is used, the formal characterization does not explain the psychological processes involved (Lewis, 1969). There is little agreement about how, psychologically speaking, we can know something together with someone else. Some authors have proposed that individuals infer common knowledge through a finite number of recursive steps (e.g., three to five, Bach & Hamish, 1979; Harder & Kock, 1976). In some of his writings, Tomasello (e.g., 2008, 2011) takes a modified version of this approach, claiming that some type of recursive mind-reading is involved in joint attention, even in infants. However, others have pointed out that the processing demands for even just a

few steps of recursive thinking are high, thus casting serious doubt on whether this can be how we – even as adults – achieve common knowledge instantaneously and effortlessly in most everyday circumstances (e.g., Clark & Marshall, 1981; Sperber & Wilson, 1986). Clark (1992) has suggested that we represent common knowledge as a 'single mental entity', e.g., "You and I mutually know X". Similarly, Pinker (2007) proposed that we employ a simplified recursive formula, such as Y: "You and I know X, and you and I know Y", where Y is both the entire formula and part of it.

More recently, there is growing momentum in the idea that there is a crucial difference between the social-cognitive processes that take place when we act as detached observers of others (adopting a third-person perspective) versus when we directly engage in interaction with them (adopting a second-person perspective). It has been claimed that the primary way of understanding others is not through observing but rather through interacting and experiencing with them; consequently, there are calls to increase scientific attention to second-person engagement (e.g., DeJaegher, DiPaolo, & Gallagher, 2010; Gallotti & Frith, 2013; Heal, 2005; Michael, 2011; Reddy & Morris, 2004; Reddy, 1996; Schilbach, 2010; Schilbach et al., 2013; Zahavi, 2008, 2011). Regarding joint attention more specifically, it has been suggested that a second-person relation between individuals (Eilan, submitted; Gómez, 2005) and intersubjective engagement (R. P. Hobson, 2005; Reddy, 2011) are necessary conditions for joint attention. These ideas have direct implications for how the jointness and common knowledge in joint attention (and joint action) could be achieved.

Therefore, here we develop these ideas further by taking a new look at joint attention and common knowledge and systematically distinguishing several closely related but clearly distinct phenomena within them. We begin by outlining a typology of different social attention levels,

which are all currently referred to in the literature as joint attention. We call them 'social attention' because all levels include consideration of another person's attention; thus all levels depict some form of triadic relation between self, other, and an object of their attention. Along with these levels, we propose corresponding distinctions for the social knowledge states that are involved in these attention levels. We then introduce the idea that the social attention and knowledge levels exist on a 'scale of jointness' (i.e., in terms of how much the other is in mind, the certainty of jointness, and the felt connection with others), and describe several factors (e.g., goals, saliency, timing and contingency, behaviors) that can cause shifts in this scale. Next, we discuss the complexity of social attention and some of the different functions of these levels (supporting coordination, heightening and creating obligations, social bonding). Finally, we suggest directions for future studies: We briefly sketch how the levels map onto the available findings in human infants and children, animals, and social robots, to highlight gaps in our knowledge, and show how the typology can be applied to other joint mental states.

2. A typology of social attention and social knowledge

First, a few general notes. Although the literature on joint attention has primarily focused on the visual modality, note that each of these levels can apply to all types of external sensory inputs (visual, tactile, auditory, olfactory, and gustatory e.g., Bigelow, 2003); that is, not only can we look at something together, we can also, for example, listen to or smell something together. Moreover, objects of social attention can be present, past, future, or imaginary events and states (including the lack of something); the self or the other (Reddy, 2003); and even mental states such as thoughts, beliefs, ideas, plans, etc. (For example, two people are discussing their plan to go for a picnic. During their discussion they are attending not to a physical object but to the idea of going for a picnic; O'Madagain & Tomasello, in press.) In addition, social attention

can come about through both bottom-up processes (i.e., through automatic, reflexive shifts of attention to a salient stimulus) and top-down processes (i.e., in an active, goal-directed way; Brinck, 2004; Kaplan & Hafner, 2006). Finally, for the sake of simplicity, most of the time we talk about two individuals, but of course more individuals can engage in each level as well. We use examples involving humans, but our framework does not presuppose that only humans are able to engage in these levels (see below for more on this).

In what follows, we define four levels of social attention (monitoring, common, mutual, and shared). We categorize the levels according to the participants' perspective (i.e., detached, third-person perspective vs. engaged second-person perspective), the type of knowledge they have when they are in these attentional states (individual vs. 'common knowledge' of different types), the participants' dependency on each other (independence vs. interdependence), and the kind of experience the participants have (individual vs. co-created), and we discuss some of the behaviors involved in each level. Above we used the term 'common knowledge' broadly, as it is currently used in the literature, to refer generally to states that involve 'knowing together', as it is a well-established term. However, in what follows, we further distinguish common, mutual, and shared knowledge to refer to the specific knowledge states associated with the corresponding attention levels.

A necessary pre-condition for each of the four social attention levels is that one is able to engage in *individual attention.* Individual attention simply means attending to something while engaging with the environment from a first-person perspective only. The attender is completely independent of others (no others need even be around), she has *individual knowledge* about the object of attention, and she does not take the perspective of anyone else present or change or connect with their attention state.

2.1. Monitoring attention

An individual is *monitoring attention* when she takes a third-person, observer's perspective on a second individual, and attends to what the other is attending to.

Example 1: On an airplane, Allison notices that Rob, who is sitting next to her, suddenly moves his head closer to the airplane window and looks out. Allison follows Rob's gaze and sees another airplane passing by and, at the same time, knows that Rob sees the airplane as well. Example 2: Annie is attending an anti-government demonstration. Suddenly, the sound of a loud siren fills the area. To protect her hearing, Annie covers her ears with her hands. A stranger, Alvy, stands in front of Annie with his back turned to her. Annie sees that Alvy also covers his ears when the siren starts. Annie is paying attention to the siren and, at the same time, focuses on the fact that Alvy is paying attention to the siren as well. Annie has *individual knowledge* about the siren and about the fact that Alvy perceives the siren as well. Her attention level is *independent* of that of Alvy in the sense that she can engage in this level separately, without Alvy's active contribution or even awareness that she is there. Annie is having an *individual* experience. In this level, individuals do not have to be in the same level (as they do in the common, mutual, and shared levels below). Here, Annie is monitoring attention and Alvy is in individual attention. However, note that even when both individuals are monitoring each other's attention to the object simultaneously, each just assesses the attention and knowledge state of the other individually, so both their experience and their knowledge remain individual. Often, monitoring attention results in an observable change in the monitor's behavior (e.g., turning one's gaze, head, and/or body orientation to look at what the other is looking at), but it is also possible to monitor someone's attention covertly, without producing any easily observable behaviors.

One can go beyond monitoring another's attention and additionally manipulate it. An individual is *manipulating attention* when she takes a third-person, observer's perspective on a second individual, monitors the other's attention, and then acts in such a way as to get the other to attend to something else. Like monitoring attention, manipulating attention involves *individual knowledge and experience* and can be *independent* of the other in the sense that we can manipulate others' attention without even being there (e.g., if we leave something out in a conspicuous place for them to see when they get home).

One can intentionally manipulate another's attention in several ways. First, one can increase the saliency of events such that they capture the other's attention. For example, if a guest at a dinner party does not want to ask directly for more wine, she can surreptitiously move her empty glass into the host's visual field so the host will likely notice it and refill it (Sperber & Wilson, 1986). This requires monitoring what the host can/will be able to see. Second, one can directly manipulate someone else's attention via a sort of unidirectional communication. The goal of unidirectional communication is only to convey information (e.g., there is something over there) from one individual to another, not to create a common focus of attention (in contrast to the bidirectional communication that takes place in the shared attention level below). Gestures such as pointing can be used to direct someone's attention in this way. For instance, when the siren starts, Annie turns to Alvy, points ambiguously to the side while shouting, "The police are coming, look!" Alvy searches for the police. Annie uses the moment of distraction to steal Alvy's bag and run away.

In contrast to the monitoring attention level, which involves individual knowledge and experience, some type of 'common knowledge' (again in the broad sense of the term) about the

focus of the individuals' attention is an integral part of each of the next three (common, mutual, and shared) levels. That is, in each of the next levels, it is necessary that both individuals are in the same level in order to achieve these levels.

2.2. Common attention

Two individuals are in *common attention* when they each take a third-person, observer's perspective on the other, more or less simultaneously attend to what the other is attending to, and infer not only that they are both attending to the same thing, but also that they are attending to each other's attention to the thing. Analogously to the classic definition of common knowledge discussed above, this characterization lays out what is needed to achieve common attention without committing to the specific psychological processes involved (see also Geurts, in press). That is, depending on the situation, individuals might engage in higher-order recursive reasoning (up to varying numbers of orders) or use various shortcuts to infer common knowledge (e.g., Clark, 1992; Pinker, 2007). The key point here is that this is done from a third-person, observer's perspective and that both individuals believe that they are in common attention.

Individuals can engage in common attention when 1) the object of attention is salient or public (so they can each assume that they are attending to the same thing), 2) for each of them the other's attention is relevant for them (so they each have a reason to consider the other's attention, for example they are in close physical proximity, or they have a previously-established joint goal, or they want to predict each other's actions), and 3) they each know that their own attention is relevant and salient to the other. Under these conditions, individuals could know that they are attending to each other's attention without any contact or communication.

Example 1: Suddenly, the lights go off in the plane and Allison and Rob find themselves in darkness. Allison is attending to the fact that it is dark, to the fact that Rob must also be

perceiving the darkness, and to the fact that each knows that they each know this, and Rob is engaging in the same attentional processes. Example 2: Alvy stands next to Annie, so they are aware of each other's presence. Annie is attending to the siren, to the fact that Alvy is attending to the siren, and to the fact that each knows that they each know this since the siren is so loud that it is clear that everyone in the vicinity must be hearing it. Alvy is engaging in the same attentional processes. Annie and Alvy have *common knowledge* about the siren achieved via recursive assumptions, inferences, and perspective-taking, since there are no specific external behaviors (beyond monitoring attention) associated with this level. Their attention level is *dependent* on that of the other in the sense that, from a bird's-eye perspective, both of them must be attending to each other's attention to the thing and each other to achieve this level. That is, the dependency is in their awareness that they are both engaging in the same attention processes. From each individual's perspective, however, each evaluates separately whether they are in common attention and each might be mistaken about it (e.g., one believes that they are in common attention but later finds out that they were not). Therefore, Annie's and Alvy's experience, when reasoning about each other's attention, is *individual.*

2.3. The question of jointness

This last point is key in separating the common attention level from the next two (mutual and shared) levels. According to one of the richest definitions of joint attention, individuals must *know together* that they are attending to the same thing (Tomasello, 1995). It might seem that already in common attention Annie and Alvy are attending together to the loud siren, because they have common knowledge about it. However, we would argue that they are not truly attending *with* each other yet, as they have not yet made 'attention contact' with each other (Gómez, 1996), actively *shared* the information in any way yet (Carpenter & Liebal, 2011), or in

Taylor's (1985) terms, they have not yet put it out there *'between us'* as something to focus on together. Because in the common attention example Annie stands in a third-person relation to Alvy at this moment, she must infer and assess Alvy's knowledge. Using either the classic recursive approach (s/he knows that I know that s/he knows…) or some simpler version of it (Clark, 1992; Pinker, 2007), they can each know that they each know about the siren and each other's knowledge of both it and their own knowledge, and thus can achieve a type of 'common knowledge' (broadly speaking). However, this level of attention and knowledge does not seem to be known truly together in any meaningful way. In contrast to the 'meeting of minds' that takes place in the classic 'joint attentional triangle', it is basically two individual perspectives that never meet in the middle: Each individual just assesses the attention and knowledge states of the other individually (Calabi, 2008; Carpenter & Liebal, 2011).

Although fairly complex processes including recursive assumptions, inferences, and perspective-taking can be used both in third-person as well as in second-person relations, in second-person relations, engagement can allow for a different, more direct and non-inferential processing of the situation, which is not possible outside of this type of relation (Gómez, 2005; Reddy, 1996). Thus the proposal is that we should reserve the terms 'mutual' and 'shared' for interactions that involve this second-personal relation – where it is really done *together* (Eilan, submitted; Gómez, 2005; Reddy, 1996, 2011; Zahavi, 2015).

In a second-person relation, one no longer has a detached, observer's attitude toward the other and his or her attention; the other is not perceived as 'he' or 'she' (Reddy & Morris, 2004). Instead, both individuals are in direct engagement with each other and treating each other as 'you', and both are senders and receivers of information at the same time (Argyle & Cook, 1976; Zahavi, 2015). Direct social interactions provide both partners with a wealth of cues about what

is relevant and salient, and where each other's attention is focused. Importantly, each partner becomes a "constituent part" of the other's experience (Campbell, 2005; Zahavi, 2015). That is, their perspectives and attention to the object of attention are colored by their mutual awareness of each other's attention, and the experience differs qualitatively from the situation in which the individuals attend to the same thing but are not relating as 'you' to each other (as in the monitoring and common attention levels above).

Although second-person engagement is typically described as being an extremely rich and in-depth intersubjective experience, actually it comes in degrees. Even within an interactional context, we can take different approaches towards the other (i.e., instrumental, detached, and objectified vs. engaged, involved, and attuned) – not all interactions involve mutual engagement (e.g., Buber, 1958; Fuchs, 2012, Reddy, 2018; Schilbach et al., 2013). We will thus use the term second-person relation to refer to situations in which individuals adopt an engaged approach towards each other in an interactional context. Both the intensity of interactive exchanges and participants' approach influence the richness of a second-person relation. A prolonged, face-to-face conversation between two close friends represents a rich version of this, but even a brief encounter between two strangers making eye contact can establish a second-person relation, although a very minimal one, as it already allows for bidirectional contact and signals openness for engagement (in contrast we skillfully avoid eye contact when no engagement is desired.)

This is the crucial difference that we would like to highlight as we think about the difference between the previous levels and the next two: The previous levels have an individualistic nature; the next levels, mutual and shared attention, are bidirectional and relational. The main difference between the next two levels lies in the presence or absence of

intentional bidirectional communicative exchange about the object of attention.

2.4. Mutual attention

Two individuals are in *mutual attention* when they are engaged in a second-person relation to each other, while they are more or less simultaneously attending to the same thing, such that both of them directly experience each other attending both to that thing and to their own attention to the thing and each other. Thus, mutual attention involves both individuals being aware – in a second-person relation to each other, but without intentional communication – that they are attending to the same thing with each other. Most prototypically, the individuals make eye contact about the thing. When their eyes meet, both know that they are attending to each other and 'attention contact' is established (Gómez, 1996). When the thing in the environment is salient and/or relevant for both individuals, eye contact enables them to gain reciprocal and reactive information about their attention to it. Another example of such bidirectional behavior is mutual touch (Botero, 2016) (e.g., squeezing hands).

Example 1: When the lights go off in the plane, Allison and Rob grasp each other's hand reflexively. Example 2: When the siren first starts, Annie looks around to see what is happening. So does Alvy, and their eyes meet. As their eyes meet, it is mutually manifested that they are both attending to the siren and thus they both know that they both know this. Annie and Alvy have *mutual knowledge* about the siren achieved via experiencing it with each other. Their attention level is *dependent* on that of the other in the sense that both of them must engage in these attentional processes together in order to achieve this level. Their experience is qualitatively different from the individual experiences in the previous levels; they *co-create their experience* and it is colored by their direct mutual awareness of each other's attention. Even though the eye contact signals to both individuals that the contact was established, in mutual

attention neither of the individuals intentionally communicates about the object of their attention.

2.5. Shared attention

Two individuals are in *shared attention* when on top of an interaction that meets the criteria for mutual attention, in addition they intentionally communicate with each other about the object of their attention, and/or the fact that they are sharing attention to it. In contrast to mutual attention, which apparently typically comes about incidentally (e.g., via incidental eye contact), the state of shared attention is achieved intentionally and can often be an end in itself (Gómez, Sarriá, & Tamarit, 1993).

Example 1: Rob sees the other airplane passing by in the distance. He points out the window for his son sitting next to him, waits until his son looks out at the airplane, and then they engage in eye contact and start talking about the plane. Example 2: When the siren sounds, as Annie starts looking around to see what is happening, so does Alvy, and their eyes meet. As their eyes meet, Annie raises her eyebrows questioningly, and Alvy responds with a concerned look. From the communication inherent in shared attention, Annie and Alvy gain several layers of *shared knowledge* about the siren (see below). Annie and Alvy's attention level is *dependent* on that of each other in the sense that both must engage in these attentional processes in order to achieve this level. Again, they *co-create their experience*, but the experience is different from mutual attention as the communicative exchange allows them to actively align psychological states such as attitudes, emotions, and goals about the object of attention with each other.

In the shared attention level, individuals *follow* and *direct* others' attention, but in contrast to the corresponding but third-person monitoring and manipulating attention above, the goal here is to create a bidirectionally-shared focus of attention (Gilbert, 2007). Often, in the top-down case, one partner intentionally directs another's attention to something they can focus on

together using behaviors that are intentionally communicative, checks that the other has perceived it, and the other partner confirms, communicatively, that attention is shared. Shared attention is thus characterized by behaviors in which individuals confirm to each other that they are attending to the same thing. The communication need not be verbal (although often it is, e.g., "Look!"); it could also, or in addition, simply involve 'communicative' or 'sharing' looks (e.g., Carpenter & Liebal, 2011; J. A. Hobson & R. P. Hobson, 2007), that is, looks that express some message. Other prototypical examples of nonverbal behaviors are declarative gestures such as pointing and showing. This confirmation is present in the bottom-up case as well. When a stimulus is salient, partners first automatically shift their attention to it, and then they confirm verbally (e.g., "That was loud!") or nonverbally (e.g., eye contact with nodding or smiling) that they are sharing attention (Brinck, 2004; Kaplan & Hafner, 2006).

Several layers of messages can be communicated in shared attention, which add to the layers of shared knowledge that can be gained. Not only do the partners have shared knowledge about the object of attention (the main referent or topic of their communication, e.g., we hear the siren), they also have shared knowledge about the communication itself (e.g., we intentionally communicated about our attention to the siren). In addition, partners also typically convey an attitude or comment about the object of attention (Bruner, 1974; Carpenter & Liebal, 2011; e.g., we are concerned about the situation), so they have shared knowledge about this as well. Sometimes they may even comment on the fact that attention is shared, or "share the experience of sharing" (R. P. Hobson & J. A. Hobson, 2008, p. 79), for example when an infant and her mother look at each other with a smile while playing, for no other reason than to express pleasure about their shared activity. The communication inherent in this level means that all of these things are out in the open, public, and, to some extent, undeniable (Carpenter & Liebal, 2011).

All of these layers create more specific shared common ground about the object of attention, the partners' attitudes about it, and their shared experience, and contribute to building up the partners' conversational record (Lewis, 1979).

To summarize, here we have offered a new typology of different social attention and corresponding social knowledge levels, highlighting the difference between third-person and second-person relations. In contrast to the way in which the term joint attention has previously been used quite broadly in the literature, we suggest that we are only jointly attending to something with someone when we are in a *direct, reciprocal, and engaged second-person relation with them*. The most prototypical case of this is the shared attention level; the minimal version is the mutual attention level. The levels have a cumulative structure in terms of the layers of knowledge involved. Using the siren example above:

i) In all four social levels, I know that I hear the siren and that you hear the siren.

ii) In the common, mutual, and shared levels, additionally, I know that each of us knows that each of us hears the siren and that we each know that we each know all this about each other.

iii) In the mutual and shared levels, additionally, I know that we are both experiencing together our hearing the siren and both know this.

iv) In the shared level, additionally, we both know that we communicated to each other about the siren and thus that we have confirmed to each other that we hear the siren together.

See Figure 1 for a schematic depiction of each level, and Table 1 for a summary of some of the characteristics of each level.

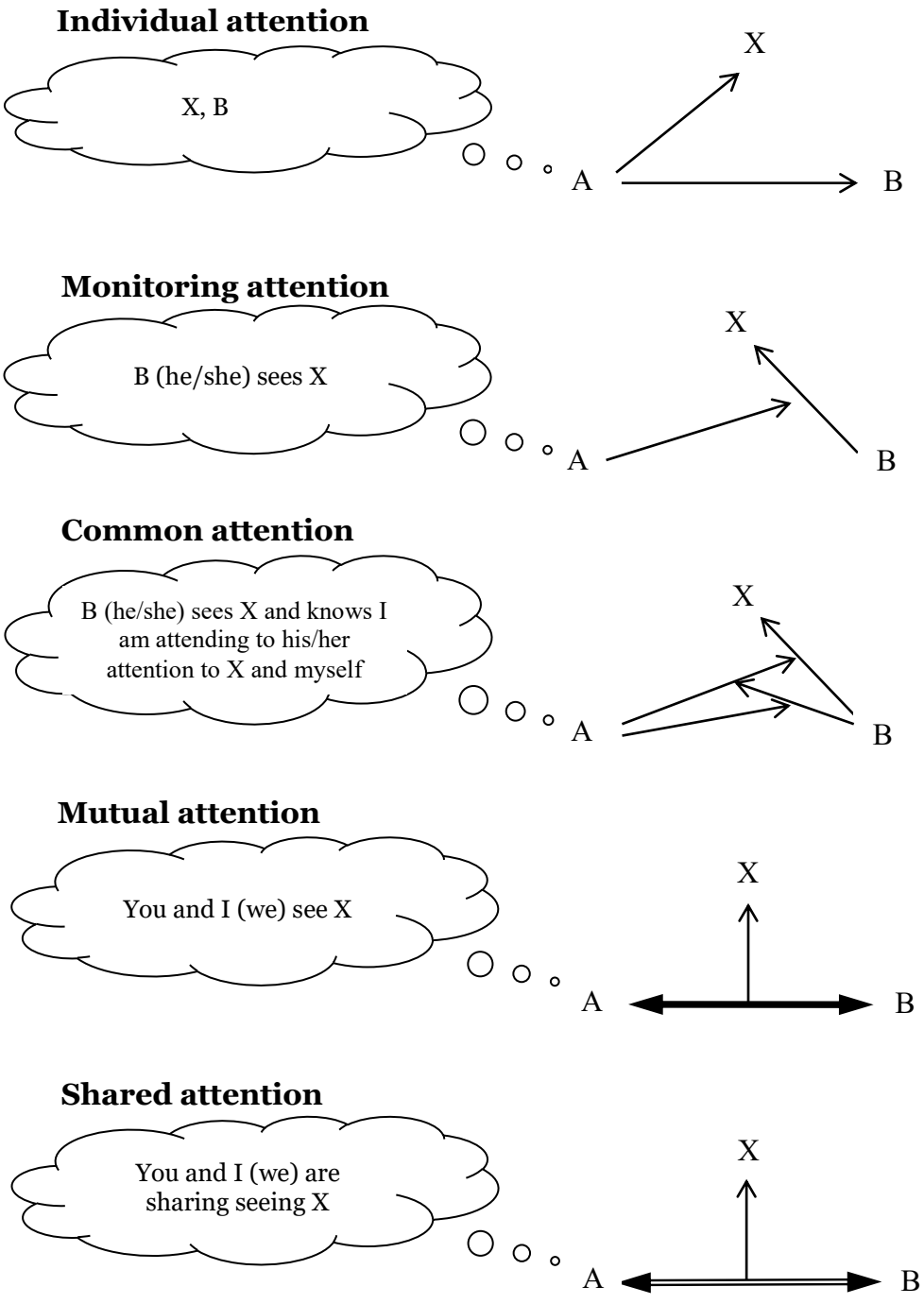

Figure 1. The figure is from the perspective of target individual A only. B is a second individual. X is the object of attention. Arrows represent what A and B are attending to. In individual attention, A attends to X and B. In monitoring attention, A attends to B's attention to X. In common attention, A attends to B's attention to X and herself. In mutual attention, A's attention to B's attention to X and herself is depicted via the solid two-way arrow, which represents non-

communicative eye contact. In shared attention, the special two-way arrow represents communicative eye contact (and/or other bidirectional communication). Thought bubbles represent what A has in mind (this figure depicts a visual example).

Table 1.

Characteristics of the different attention levels. The same characteristic across levels is marked by the same color.

	Individual	Monitoring	Common	Mutual	Shared
Type of perspective	First person	Third person	Third person	Second person	Second person
Description of the experience	I attend to X	He/she attends to X	He/she attends to X and to my attention to X and himself/herself	You and I (we) attend to X	You and I (we) attend to X and are actively communicating about this
Type of interaction	Dyadic (i.e., subject and object)	Triadic	Triadic	Triadic	Triadic
Knowledge gained	Individual knowledge	Individual knowledge	Common knowledge	Mutual knowledge	Shared knowledge
Does the second individual know that the first is in that level of attention?	Not necessarily	Not necessarily	Yes	Yes	Yes
"Intimate I+you we feeling"[1]	No	No	No	Yes	Yes, stronger
Commitments and obligations	No	No	Yes (but plausible deniability is possible)	Yes (but some plausible deniability is still possible)	Yes

[1] There are several types of "we feeling" (Zahavi, 2015). Here we distinguish a) the intimate/private/personal we feeling arising from a bidirectional, direct connection with others (this we feeling is subjectively strong) and b) the associative we feeling arising from association with others, arising because individuals have something in common (e.g., they both are attending to the same thing; this feeling comes from knowing rather than experiencing). It is interesting to note here that while in mutual and shared attention, almost by definition you know exactly who you are attending with, in common attention this does not need to be the case. For example, I can know that everyone (with normal hearing) within a certain radius must have heard the noise of a loud explosion, without knowing exactly who those people are.

3. A scale of jointness

In the literature, joint attention is discussed as a binary, discrete, and uniform event: Either social partners are in joint attention or they are not, and when they are, they are in it in an all-or-nothing way. However, we argue that in actual fact, there is much more complexity to social attention, and that jointness comes in degrees. This is the case both when comparing different levels and when considering the jointness within each level. There are various things that make situations feel more or less joint. Aspects of jointness include (but are not limited to) the perceived psychological closeness between the individuals, how much the other is in mind, the connection felt (the 'we feeling'; Zahavi, 2015), and the certainty about the object of attention and/or that attention is joint (see below for more on this latter point). That is, first, the social attention levels exist on a scale of jointness (see Figure 2), with monitoring attention (in which we have an individual experience of the other's attention) on the left side, representing no jointness, and shared attention (in which we truly attend to something together with the other) on the right side, representing the highest degree of jointness. Second, it is important to note that although for the sake of simplicity we provided prototypical examples for each level above, in reality there exist stronger and weaker examples in each level, that is, that there is a continuum within each level too. Consider the siren example in the mutual attention level. This situation could range from accidental eye contact (while Annie and Alvy are instinctively looking around), to more intentional eye contact (e.g., Annie and Alvy each look at each other to check if the other hears the siren too, and they happen to do this at the same time), to eye contact with mutual reactiveness (during the eye contact Annie and Alvy both reflexively cover their ears). These examples lie to the left, middle, and right side of the scale in the mutual attention level, respectively.

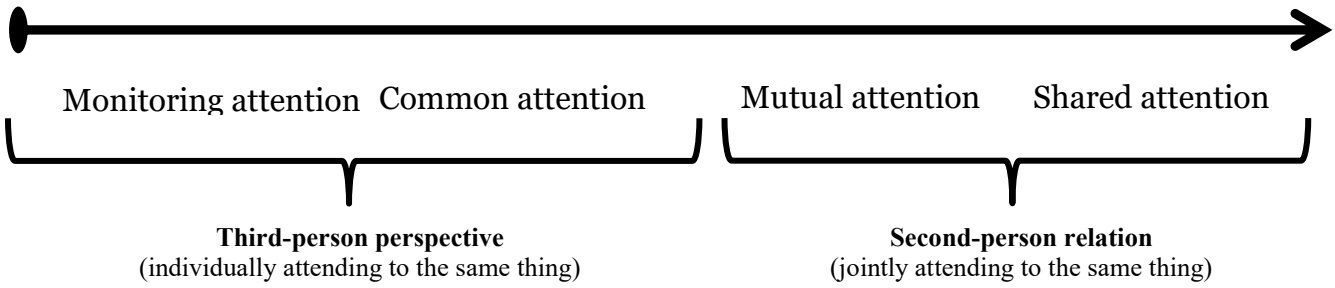

Figure 2. A scale of jointness.

3.1. Factors that can cause shifts in the scale of jointness

There are a variety of factors that can make it more likely that we move along the scale, both along the different levels of the scale and within each level:

- **Saliency:** Salient stimuli generally push situations to the right on the scale of jointness. It is easier for individuals to establish all of these types of attention to very salient things (e.g., an elephant standing in the lobby of one's office building) then to subtle things (e.g., a new vase standing in the lobby). However if a stimulus is too salient and attention-grabbing, we might focus solely on it and be stuck in individual attention (e.g., when a bomb explodes, we might at least initially just focus on ourselves and our own safety before thinking of anyone else).
- **Goals:** The existence and strength of a relevant shared goal between the individuals, and their interest in it, pushes situations to the right on the scale of jointness. If Annie and Alvy had recently spent long hours planning how to avoid being arrested at the anti-government demonstration, when the noise of the siren fills the area it is probable that they will be more on the right side of the scale compared to a situation in which they had not had a shared goal.

- **Type of common ground[2]:** Similarly, the strength and depth of the common ground the individuals have can push the situation to the right on the scale. Increasingly joint types of common ground generally range from no common ground, to common ground for rational beings, to cultural common ground, to common ground from co-presence, to personally-shared common ground (Clark, 1996; Clark & Marshall, 1981). Example: Robin and Max are waiting at a bus stop and suddenly a bomb explodes in front of them. Contrast the situation in which Max is the 2-year-old son of the woman standing next to Robin and the situation in which Max is an adult colleague who had just recently discussed with Robin a newspaper article saying that there is a security threat in their country (so, in addition to all the other types, they have personally-shared common ground). In the first case, after looking communicatively to each other, because they share little contextually-relevant common ground, Robin and Max are on the left side of the scale in shared attention about the fact that the bomb exploded. In contrast, in the second case, they have quite specific shared common ground that is relevant to the explosion, so they are on the right side of the shared attention continuum. On top of their shared attention about the bomb exploding, they are sharing a more complex and detailed message such as, "Oh, no, it is here! We talked about this yesterday." Note that if we have very specific shared common ground (especially recently shared common ground) we can communicate much more complex messages with less obvious behaviors. For

[2] Common ground suffers from terminological and conceptual confusion as well. In this paper we will use "common ground" as an umbrella term encompassing 'common knowledge' (in the broad sense of the term), beliefs, and assumptions coming from different sources (e.g., co-presence, joint personal experiences, cultural background, etc.; Clark, 1996).

example, if two employees have just been complaining about their boss, the briefest, most neutral and surreptitious glance between them during a meeting with her can convey a complex message (e.g., "See what I mean?"), especially if the look occurs contingently to something relevant that their boss just said.

- **Contingency and timing:** As suggested above, strong contingency and/or a short time interval can push situations to the right on the scale (e.g., in the shared level, contingency and a short time interval between the appearance of the object of attention and the eye contact/communication, or between the formation or most recent discussion of a shared goal and the appearance of the object of attention). For example, in the siren example, consider the difference between cases in which Annie looks communicatively at Alvy immediately after the siren starts, 10 seconds later, or 10 minutes later. An immediate look would create a stronger feeling of jointness (lying more on the right side of the shared attention level) than a delayed look, and if they looked at each other 10 minutes later, it might not be at all clear what the other's look was about.
- **Perceptual space:** Perceptual space with clear borders (an office, a department's mailing list) or small perceptual spaces (a car) push situations to the right on the scale (especially with regard to the certainty that we are attending to the same thing with someone) compared to perceptual spaces with vague borders (a university, Facebook) or huge perceptual spaces (a city, a concert hall).[3]
- **Behaviors:** Observable behavior, ranging from unintentional or reflexive behaviors to

[3] There is an interplay between the type of perceptual space and the type and saliency of the stimulus. For example, it is difficult to guess where the borders are for being able to perceive sound from an explosion, but it is easy to guess that the smell of freshly-made cookies in a room will not reach the space outside the room when the door and windows are closed.

verbal or otherwise explicit communication, pushes situations to the right on the scale. For example, when someone sneezes, it might send a signal to others that he can smell the pepper spilled on the table, or when someone gasps or sighs while watching a movie, it might send a signal to others that he was attending to something in the movie. And of course in the shared attention level, sharedness becomes more clear depending on the explicitness and ostensiveness of the communication. To illustrate, consider the example with the siren. As Annie's and Alvy's eyes meet, if they look silently at each other with communicative looks after the siren filled the area they would be more to the left of the continuum in the shared attention level than if Annie looked to Alvy and said, "Oh, this is bad!" and Alvy agreed.

- **Individual differences:** People who tend to have others in mind more often or who have a higher motivation to share appear on the right side of the scale more often than people who do not. One reason for this is simply because they initiate shared attention more often.
- **Relationship closeness:** People who are close to each other end up on the right side of the scale more often. There is a natural tendency to have close others in mind and share attention more with people who are close (family members, friends, partners).

3.2. Certainty

Two kinds of certainty are important when attending socially with someone: 1) certainty about the object of attention – how can we know with certainty that we are attending *to the same thing* as someone else? and 2) certainty that attention is joint – how can we know with certainty that we are attending to the same thing *with someone else*? Several ways of gathering information about other people's attentional focus exist in different situations and influence our

certainty. Some require fewer assumptions than others, for example you can see your friend seeing X (gaze following), but you can also just hear (but not see) that your friend is co-present, or just assume that your friend is co-present, and therefore you can assume that your friend perceives X. Similarly, fewer assumptions are required when we can directly perceive X in the environment (e.g., we can see/smell/hear X) (perceptual attention) than when we must infer the relevant object of attention (conceptual attention, e.g., to absent referents, for instance, an infant pointing about a missing puzzle piece for her mother).

When the stimulus is very salient, the certainty that you are attending *to the same thing* as someone else does not necessarily change with the different social attention levels. We are sensitive to perceptual co-presence with other people, and expect similar perceptual or attentional abilities from them, unless we learn otherwise (e.g., we find out that someone is visually- or hearing-impaired; Clark, 1996). Therefore, when the noise of a loud siren fills the area and you can hear it, you know that everyone else in the vicinity must be hearing it too.

In contrast, being in the higher attention levels increases the certainty that we are attending to the same thing *with someone.* For example, imagine that you are sitting next to your friend in a train, and, in front of you, you see that, as soon as the woman sitting opposite you leaves for a moment, the man sitting next to her reaches into her purse and steals her wallet. Even before you and your friend engage in any behavior toward each other, you infer that the stealing is salient enough to grab the attention of your friend as well, and that your friend realizes the same about you, so you have *common knowledge* about the thief (common attention). However, your certainty that you are attending to this together increases considerably once you engage in eye contact after seeing it (mutual attention), and even more once you exchange concerned looks and/or discuss it (shared attention).

Note that it is often impossible to be 100% certain that you are attending to the same thing with someone. However, this is not a problem, as certainty is probabilistic. That is, you each believe that you are attending to the same thing together, and that the other believes this with some probability too. How high the probability needs to be depends on the current purposes. So although we use terms like 'knowing together' and common, mutual, and shared knowledge, more specifically we mean probabilistic knowledge/belief with sufficiently high probability for those purposes (Clark, 1996; Monderer & Samet, 1989). To illustrate: If you are considering confronting the thief together with your friend, certainty is important and thus you likely would not want to rely only on common attention, but instead would engage in shared attention.

4. The complexity of social attention

Joint attention is often discussed in the literature as an isolated, momentary occurrence, in which individuals attend to a single object or event. However, many everyday social attention episodes are much more complex. First, most of the levels (i.e., all except mutual attention) can be quite extended in time. For example, a spy could monitor the attention of someone all day, and we can be in common or shared attention with other audience members at a concert or movie for hours.

Second, in any given situation, we can move from individual attention to any other level, or move among the different social levels, instantaneously. Sometimes one level might function as a precursor to another level (e.g., when you want to show a hot air balloon to your partner, you might first monitor your partner's attention, to determine where your partner's focus is, and then choose the best way to initiate attention sharing). Sometimes we can even engage in two or more levels simultaneously (e.g., if you and I are monitoring the attention of someone else

together, or when I take a third-person perspective on you during a second-person interaction, thinking, “Why is she looking at me that way?”; Reddy, 2018). This can result in a number of hierarchically-embedded states of attention occurring at the same time, either with the same or with different people. At a movie, for instance, I can simultaneously monitor the attention of a character on the screen, be in common attention with other audience members that the character is looking at something, and turn to share attention to the thing the character is looking at with my partner. Similarly, the whole ‘going to the movie’ experience is a shared attention episode with my partner, and there are numerous hierarchically-embedded, smaller shared or mutual attention episodes within that experience, when we, for example, agree that the popcorn is very salty, or squeeze each other’s hands reflexively at a scary moment, etc. Thus, although we mainly discuss the different social attention levels separately in order to highlight the important differences across them, we note that complex relations can also exist among and within them.

Third, the attention levels take place live, in real time. When we exchange letters with someone, although this is reciprocal communication, it is not shared attention (though real-time, immediate texting, in contrast, may be). However, we do create shared knowledge and common ground from these types of exchanges. We can also engage in virtual sharing in which we just imagine sharing attention with someone, for example, when I celebrate my country’s team’s goal while watching the World Cup abroad and imagine my dad cheering ‘with’ me in our home country. Again this is not real shared attention; but it can result in an amplified experience and influence future interactions and bonding between individuals. Thus, direct, real-time experience of social attention has unique qualities that cannot fully be attained with experiences that are sequential or delayed in time or space.

5. Functions of the different levels

It is useful to distinguish these levels for several reasons. First, as mentioned above, it should help in eliminating unnecessary confusion and debate in discussions about whether, for example, different animals can engage in joint attention (see below for our take on this). In addition, as we will discuss here, the different social attention levels serve different functions.

Monitoring attention brings some advantages to individuals who are capable of it, over and above individual attention. First, it can be useful simply for gathering more information from the environment than one could on one's own, for instance when as a result of monitoring someone else's attention we discover something interesting or useful or dangerous that we might not have discovered on our own (e.g., Max is walking on a pier at a lake and he sees a couple in front of him excitedly peering into the water. He approaches to find out what they are looking at and sees an alligator swimming there). Second, monitoring attention is useful in making informed predictions about others' future actions, so we can adjust our own actions accordingly. This is often useful in competitive situations, because then the information gathered can be used exploitatively to gain an advantage over others (Tomasello, 2008). But we can also monitor others' attention in prosocial situations (e.g., Pam is in a store and sees an elderly woman looking up, fixating her gaze frustratedly on something on a high shelf. Pam determines that the woman is looking at a box of tea, and so approaches and gets the box for the woman). Similarly, we can manipulate the attention of others to things in the environment either to reach our own goals (e.g., to get someone to attend to something we need, for example, our wineglass that needs refilling; or by distracting their attention deceptively) or to help others to reach their goals (e.g., by redirecting their attention to something relevant for them).

Common attention similarly helps us make informed predictions about others' future

actions. However in common attention one can understand not only what someone else attends to/knows but also what the other attends to/knows about one's own attention/knowledge, and this allows for a considerably more complex appraisal of the situation. This can be especially helpful in solving 'coordination dilemmas' – situations in which it is in our interest to decide what to do based on our expectation of how our social partner is likely to behave, and our partner's decision likewise depends on what he or she expects from us (Clark, 1996; Chwe, 2013). For example, imagine that a group of protesters has agreed before going to a demonstration that, no matter where they are in the crowd, whenever they hear a siren, they will start shouting anti-police slogans. Although all the protesters want to participate as a group, no one wants to engage in this potentially dangerous act alone. When a siren sounds, the protesters' ability to engage in common attention with their fellow protesters to the sound of the siren may give them confidence to make the risky decision to start shouting. Being able to engage in common attention enables us to coordinate even in cases in which we are apart and contact or communication is not possible. However, in the common attention level we can only coordinate attention towards salient or relevant things or events. If it is not clear that everyone can hear the siren, it would be more difficult or even impossible to engage in common attention with any certainty.

In contrast to the individualistic nature of these two levels, in which we must examine the situation and consider our partner's attentional state from the perspective of an outside observer, the direct nature of the attention contact in the mutual and especially shared attention levels, as well as the bidirectional, reciprocal evidence it provides about where each partner's attention is focused, makes communication, interaction, coordination, and collaboration much more efficient and fluent. Another helpful function of shared attention is that it enables us to create a shared

focus even to subtle things in complex environments, which might not be perceived as salient and/or relevant for both of us at first. This is not only helpful in supporting conversations and joint action, but in learning and other contexts as well (e.g., for infant language acquisition).

Another important function of the social attention levels is that some of them heighten or even create social obligations and commitments. For example, imagine that Jane slips on the wet floor and is having difficulty standing up again. A man is standing nearby. We all know that we should help others when we can, but sometimes we shirk our obligation to help. However, if Jane's fall makes so much noise that she and the man both know that they both know about it, then this amplifies the already-existing obligation of the man to help Jane because they are in common attention about her need for help. If this turns into mutual attention, when Jane and the man make eye contact, this feeling of obligation becomes much more personal and directed, so the man might feel particularly obligated to help her. In this case, it would be difficult for the man to deny having seen Jane fall, because the same single event of making eye contact when seeing Jane's fall also removed any doubt in the man's mind that she saw him seeing her fall. Consequently, mutual attention can create some pressure to follow social norms in situations in which people need to decide whether to act in accordance with norms or not.[4]

However, in addition to cuing or heightening existing obligations even more strongly, in the shared attention level, new obligations and even commitments can be created. The bidirectional communication in shared attention makes public the fact that attention is shared. We can refer to the instance of sharing afterwards and it is difficult to plausibly deny both the fact that attention was shared and any commitments or obligations following from this. This

[4] We thank Sebastian Grueneisen (personal communication) for the examples highlighting how common, mutual, and shared knowledge heighten the pressure to follow social norms.

facilitates social coordination above and beyond common and mutual attention. For example, in the right context, shared attention in the form of communicative eye contact can lead not only to an expectation that one's partner will cooperate, but even to the perception of a commitment on her part to cooperate (Siposova, Tomasello, & Carpenter, 2018). This perception of a commitment is so strong that it even engenders normative protests if the partner does not subsequently behave as expected.

Social attention is not only useful, but it can be a source of pleasure for us as well. This is especially pronounced for the shared attention level. It feels good to connect with others and to engage in shared attention (Schachter, 1959); we often engage in shared attention just for the sake of the sharing itself (Gómez et al., 1993). Even as 1-year-old infants, we are so motivated to share attention with others that we will actually turn away from interesting sights to share them with others − the sharing is just as pleasurable (or even more so) than the sight itself. We have a feeling of 'contact' with the other, and consequently we may feel a "we feeling" which causes changes in our experience. In shared attention, we build our shared common ground. We know what objects and events we have shared, and can refer to them in the future. Sharing attention might also cause an increased sense of similarity (with all the benefits that that entails; Haun & Over, 2013), which may emerge because paying attention to the same things/events/ideas might lead to thinking that whoever is interested in the same things I am is probably like me. Similarly, both initiation of shared attention and responding to shared attention signal an interest in the other – an interest in exploring the world together. With shared attention we can also express trust, openness, and intimacy. There are individual differences in the ease with which we share attention with others. The closeness of our relationship with our partner may increase the probability of smooth, mistake-free episodes, and, in turn, successful episodes of shared attention

likely increase bonding (see also Wolf, Launay, & Dunbar, 2015). Thus, shared attention is important in forming attachments and sustaining positive interpersonal relationships. On the other hand, skilled deceivers can exploit this to create an illusion of openness and use shared attention behaviors for deception in competitive or deceptive situations (e.g., being sure to make eye contact when lying).

6. Future directions

The next important step in social attention research is to come up with operational definitions of these levels so it will be possible to determine which levels different types of individuals, such as young children, animals, and social robots, can achieve. As social attention states cannot typically be measured directly (except perhaps using neuroimaging techniques), it is challenging to find the best ways of investigating them.

One approach is to look for specific behaviors. The question is which behavior(s) should be considered as indicative of each of the levels. We have described some behaviors that typically accompany the levels above. However, these particular behaviors might differ across species and are neither necessary nor sufficient for identifying any given level. For example, turning one's head/gaze in the direction of someone else's could be a result either of low-level spatial cuing or monitoring or following someone else's attention. And a quick glance between partners could either be an accidental meeting of the eyes or sharing attention via communicative but neutral eye contact (e.g., in a situation, such as in a group meeting, in which the signaler does not wish anyone other than the recipient to be able to discern his meaning). It might thus be helpful to focus on behavioral sequences, instead of isolated behaviors. An example of this could be looking at partners' interaction history: for example, what preceded the quick glance between them, whether the glance was contingent on something salient for them, and what behaviors

followed afterwards. However, quite often we do not have access to others' interaction history or future interactions. And although useful, investigation only of observable behaviors without considering knowledge states cannot provide conclusive evidence about which levels different types of individuals can achieve.

Experiments are helpful in this regard. For example, one approach is to design studies in such a way that participants can only succeed on the task if they are engaging in a certain attention and knowledge level, or to experimentally manipulate the level that the participants can be in. It could help, for example, to focus on individuals' behavior in situations such as coordination dilemmas, in which their patterns of responses are predicted by their attention and knowledge states. For instance, in a Stag Hunt coordination game, in a common attention condition, no eye contact or communication with one's partner could be possible, but the presence of the stag could be made salient and public in other ways, whereas in mutual and shared attention conditions, eye contact and/or communication could be possible (for an example of a similar approach, with different patterns of responses in different conditions, see Thomas, DeScioli, Haque, & Pinker, 2014). Another useful approach is to test whether participants can engage in a given level and include control conditions designed to rule out lower-level alternative explanations of participants' behavior (see, e.g., studies on infants' shared attention by Liszkowski, Carpenter, Henning, Striano, & Tomasello, 2004, and Moll, Richter, Carpenter, & Tomasello, 2008). Still another approach is to experimentally manipulate characteristic features of the different levels in order to empirically test claims about, for example, the role of communication in achieving shared attention (e.g., see Siposova et al., 2018, for an example of this approach).

An important direction for future research is to determine which levels various types of

individuals show. It is beyond the scope of this paper to review this in depth, but next we sketch briefly how the proposed levels map onto the available findings in human children, animals, and social robots. Note that as previous studies were not designed to distinguish among these levels, at this point it is not always possible to make firm conclusions about which levels different individuals can achieve.

6.1. Which attention levels can different types of individuals achieve?

6.1.1. Human infants and children

A pre-condition for each of the four social attention levels is some understanding of attention, or at least perception. For example, minimally, it is necessary that individuals understand that when others are 'looking at something,' they are 'seeing something.' It has been proposed that already in their first months infants demonstrate an awareness of others' attention by producing emotional reactions when experiencing others' attention to themselves (Reddy, 2003), and by decreasing smiling after interrupted eye contact at the age of 3 months (Hains & Muir, 1996). Even newborns look preferentially at faces with direct gaze (Farroni, Csibra, Simion, & Johnson, 2002) and perceived gaze shifts trigger corresponding gaze shifts in 3-month-olds (Hood, Willen, & Driver, 1998). However, it is not clear how much infants understand about others' attention at these young ages. It becomes clearer around 12 months of age that infants do have some understanding of others as capable of attending to things, when they follow the gaze of others to things that they cannot see themselves, for example around visual barriers and even to things behind them (e.g., Deák et al., 2000; Moll & Tomasello, 2004).

Although we presented the levels in the order monitoring, common, mutual, and shared attention (representing the degree of jointness involved), interestingly, this does not necessarily correspond with the order in which these levels emerge in human development. In fact, it appears

that the latter two levels are seen earliest in development. Most researchers agree that joint attention develops between 9-12 months of age; however others have claimed that the beginnings of joint attention actually are seen well before this. For example, Rossmanith, Costall, Reichelt, López, and Reddy (2014) have suggested that during a picture book activity, 3-month-old infants already coordinate their attention between the book and their caregiver, and that by 6 months they jointly communicate affectively about the activity. Similarly, Striano and Bertin (2005) claimed that 5-month-olds engage in joint attention after observing that they alternate gaze between an object and a partner's face. However, without any other supporting evidence of mutual or shared knowledge, the behavior of gaze alternation alone could be evidence of any of the levels: alternating individual attention between the object and the partner's face, monitoring the partner's attention, or common, mutual, or shared attention. So far, the earliest supporting evidence of shared attention with shared knowledge from experimental studies is with 1-year-old infants. Starting from at least 14 months of age, infants distinguish what objects they have and have not shared communicatively with their partner (Moll et al., 2008; see also Tomasello & Haberl, 2003) and use this knowledge both when pointing for others and when interpreting others' pointing gestures (Ganea & Saylor, 2007; Liebal, Behne, Carpenter, & Tomasello, 2009; Liebal, Carpenter, & Tomasello, 2010; Saylor & Ganea, 2007). These experimental findings, and experimental findings showing that infants point to share interest (Liszkowski et al., 2004), together with the many observational findings of spontaneous production of declarative gestures beginning between 9-12 months (e.g., Carpenter, Nagell, & Tomasello, 1998), provide strong evidence for the shared attention level. As previous studies have not distinguished between mutual and shared attention, whether these levels develop simultaneously is an open question for future research.

In contrast, so far, the earliest experimental evidence of monitoring attention is with 18-month-old infants. For example, 18-month-olds, but not 14-month-olds, can distinguish which objects an experimenter knew about when they observed, from a third-person perspective, the experimenter manipulating the objects (Moll & Tomasello, 2007b). At least from 18 months of age children also can learn novel words when they overhear an experimenter using them with someone else (Akhtar, Jipson, & Callanan, 2001; Floor & Akhtar, 2006). So far there is little if any evidence of young children manipulating attention in a third-person context (see Grosse, Scott-Phillips, & Tomasello, 2013, for the study that comes the closest to this). It might sound surprising that the monitoring and manipulating attention levels may emerge relatively late in development, as many studies have found the parallel abilities of following and directing others' attention in younger infants, but almost all of these studies were conducted in a second-person context. Thus it is not yet known exactly when the skills to monitor and manipulate others' attention from a third-person perspective (that is, outside of a mutual or shared attention episode and without the active contribution of their partner) first emerge in infants.

Similarly, common attention has not yet been directly investigated in infants – again the vast majority of infant studies take place within a second-person context. Several findings suggest that adults engage in the common attention level, and that common attention has an impact on adults' experience of the thing they are attending to, their relationship with their partner, and their ability to coordinate in social dilemmas (Boothby, Clark, & Bargh, 2014; Boothby, Smith, Clark, & Bargh, 2016; Thomas et al., 2014; Wolf et al., 2015). With young children, the closest study so far is a coordination study by Grueneisen, Wyman, and Tomasello (2015). This study investigated young children's ability to choose the same of several options as a partner without being able to communicate with or see one another. Five-year-old and older

children, but not 3-year-olds, successfully chose the same option. To solve the task, each child had to attend to the fact that one option was more salient than the others, that the other child was likely to perceive the same option as salient too, and to the fact that each knew that they each knew this.

Thus we suggest that infants develop mutual and shared attention before they are able to understand others' attention via observing and inferring it solely from a third-person perspective without any direct contact, although the claim that monitoring and common attention might develop later needs to be confirmed with future research. Several findings support the idea that the second-person relation is the primary relation in which infants develop their earliest social attention skills (Moll & Tomasello, 2007a; Peacocke, 2005; Reddy & Morris, 2004; Vygotsky, 1978). Moll and colleagues have directly contrasted infants' ability to consider others' attention in a second- vs. a third-person relation (Moll, Carpenter, & Tomasello, 2007; Moll & Tomasello, 2007b). They found that although 18-month-olds were able to do this in both situations, 14-month-olds recognized what their partner was attending to only when they were engaged with her in a second-person interaction, not when they were only observing her from a third-person perspective. Similarly, even within a second-person relation, infants show improved social attention skills in interactions with higher degrees of second-person engagement. For example, infants are much more likely to follow gaze shifts after their partner engages in eye contact (Farroni, Mansfield, Lai, & Johnson, 2003; Senju & Csibra, 2008) or points for them (Deák et al., 2000) compared to only seeing their partner's gaze shifts or head turn.

The crucial role of social attention in interactions is especially evident when social attention skills are impoverished. Children with autism show both quantitative and qualitative differences in many social attention skills compared to typically-developing children (for a

review, see Bruinsma, Koegel, & Koegel, 2004). For example, children with autism engage less in spontaneous gaze following, although when asked, they can answer correctly which object another person is looking at (Leekam, Baron-Cohen, Perrett, Milders, & Brown, 1997). They are also less likely than other children to combine their vocalizations and gestures with eye contact (e.g., Phillips, Gómez, Baron-Cohen, Laá, & Rivière, 1995; Warreyn, Roeyers, Wetswinkel, & Groote, 2006) and engage in shared attention (e.g., Mundy, Sigman, & Kassari, 1990). Based on a series of experimental studies, R. P. Hobson (1993, 2005, 2007, 2014) has argued that these atypical behaviors are a result of deficits in second-person engagement, specifically, in difficulties with experiencing others' attitudes and emotions directly in interaction and thus identifying with one's partner's perspective. Further studies are needed to understand what children with autism know about others' attention. Especially useful would be experiments investigating their ability to achieve shared knowledge.

6.1.2. Nonhuman animals

Again, to be able to engage in social attention, a pre-requisite is to have some understanding of others' attention/perception. Numerous animal species show different reactions when others are looking towards vs. away from them, including lizards (Hennig, 1977), snakes (Burghardt & Greene, 1988), birds (Ristau, 1991), horses (Proops & McComb, 2009), dogs (Virányi, Topál, Gácsi, Miklósi, & Csányi, 2004), and many primates (for a review, see Emery, 2000). Many species are also able to follow the gaze (or at least head) direction of others, for example tortoises (Wilkinson, Mandl, Bugnyar, & Huber, 2010), goats (Kaminski, Riedel, Call, & Tomasello, 2005), dolphins (Pack & Herman, 2006), horses (Proops & McComb, 2009), dogs (Miklósi, Polgárdi, Topál, & Csányi, 1998), and again, many primates (for a review, see Emery, 2000). However, the degree to which some of these species understand others as attending to

things remains an open question. Experiments testing animals' ability to follow gaze to locations outside their own visual field provide clearer evidence that they have some understanding of others as capable of attending to things. For example, it has been shown that ravens (Bugnyar, Stowe, & Heinrich, 2004), wolves (Range & Virányi, 2011), dogs (Met, Miklósi, & Lakatos, 2014), monkeys (Amici, Aureli, Visalberghi, & Call, 2009) and all ape species (Bräuer, Call, & Tomasello, 2005) follow gaze behind barriers (see also further experiments on understanding of seeing, and knowing, including Bräuer, Call, & Tomasello, 2007; Flombaum & Santos, 2005; Kaminski, Call, & Tomasello, 2008; Kaminski, Pitsch, & Tomasello, 2013; Melis, Call, & Tomasello, 2006; Santos, Nissen, & Ferrugia, 2006).

Whether animals can engage in joint attention is a hotly-debated question. Most discussions focus on apes: Several researchers have argued that apes do engage in joint attention (e.g., Leavens & Bard, 2011; Leavens & Racine, 2009; Tanner & Byrne, 2010), whereas others have argued that they do not (e.g., Carpenter & Call, 2013; Tomasello, 2006). Evidence taken in support of the idea that apes can engage in joint attention includes findings that gorillas (Tanner & Byrne, 2010) and infant chimpanzees (Bard, Bakeman, Boysen, & Leavens, 2014) alternate gaze and gesture in triadic situations. Leavens and colleagues have also claimed that apes point declaratively (e.g., Leavens, Hopkins, & Bard, 2005; Leavens & Racine, 2009). In contrast, other studies have found little evidence of joint attention in apes, either when operationalizing joint attention as gaze alternations or declarative gestures (e.g., Bard & Vauclair, 1984; Carpenter, Tomasello, & Savage-Rumbaugh, 1995; Rivas, 2005; Tomasello & Carpenter, 2005; Tomonaga et al., 2004).

However, as noted above for human infants, gaze alternation is not in itself diagnostic of any of the social attention levels. Furthermore, the few reported 'declarative' gestures in apes

were either elicited by human caretakers (who told the apes to show them something) or were actually other, different types of gestures (e.g., informative, rather than declarative, gestures; Carpenter & Call, 2013). Although it is possible that informative and other types of gestures can involve joint attention, they need not do so (whereas, in contrast, declarative gestures provide convincing evidence of joint attention because their sole purpose is to share attention and interest – the sharing is an end in itself rather than a means to some other end; Gómez et al., 1993). More experimental research is needed to determine what apes' motivation is when gesturing, and whether, when communicating, they create a shared focus of attention (with shared knowledge) or whether instead they engage in a more unidirectional type of communication: manipulating, rather than directing or sharing, others' attention.

Future studies, along the lines of experiments conducted with human infants (e.g., Moll et al., 2008), are also needed to investigate apes' ability to have mutual and shared knowledge with their partners (whether they be conspecifics or humans). Several studies claim that apes have at least the prerequisites for an understanding of common ground, that is, that they track the knowledge states of other individuals and use them in their communication with them (e.g., Bohn, Call, & Tomasello, 2016; Crockford, Wittig, Mundry, & Zuberbühler, 2012; Genty, Neumann, & Zuberbühler, 2015). However, as the authors of these studies admit, further work is needed to distinguish whether in these studies apes were using individual knowledge about their partner's knowledge (e.g., s/he knows/does not know this – monitoring and manipulating attention) or shared knowledge (e.g., we know this together – sharing and directing attention).

Another useful approach is to conduct experiments in which participants can only succeed on the task if they are engaging in a certain attention and knowledge level. This approach was used in a study by Duguid and colleagues (Duguid, Wyman, Bullinger, Herfurth-

Majstorovic, & Tomasello, 2014), and they found that unlike 4-year-old children, who spontaneously shared attention communicatively with their partner and thus solved a coordination problem, apes did not communicate before their decisions, and therefore were less successful.

Finally, with regard to social attention, it is an open question whether and to what degree apes (and other animals) adopt instrumental vs. engaged approaches towards their partner during interactions (Buber, 1958; Fuchs, 2012; Reddy, 2018; Schilbach et al., 2013; compare with Gómez's, 2005, distinction between different notions of others as subjects). Answering this question would help us to distinguish whether, when animals gesture, they are manipulating others' attention or directing others' attention to share it. Similarly, answering this question would also help us distinguish whether, when animals follow others' gaze (or track others' attention in other ways), they are monitoring attention or following attention.

Common attention has not yet been studied in other animals. Future research could adapt previous studies investigating what apes know about others' knowledge (e.g., Hare, Call, Agnetta, & Tomasello, 2000; Kaminski et al., 2008) to include the additional step of testing what animals know about what others know about what they themselves know.

6.1.3. Social robots

Social attention abilities are critical in social robotics as well. Social robots have a rudimentary sense of what others can see and can track others' attention in simplified environments (Lemaignan, Garcia, Jacq, & Dillenbourg, 2016). So there is some limited evidence for skills of monitoring attention. Robots' subtle gaze shifts towards objects have been shown to influence human participants' attention to objects even when they were unaware of the gaze shifts (Mutlu, Yamaoka, Kanda, Ishiguro, & Hagita, 2009). This finding provides some

evidence that robots could be used to manipulate others' attention. The use of social gaze is widely studied in social robotics, as it is a key component in regulating human-robot interactions. For example, robots can use direct gaze to initiate a conversation, use gaze aversion to facilitate turn-taking, and combine gaze with gestures and speech (for a review, see Admoni & Scassellati, 2017). Their repertoire can include intentionally communicative behaviors that are designed to evoke a feeling of shared attention in their human interaction partners, such as declarative pointing and vocalizations ("Look!"; Anzalone et al., 2014). Therefore, the behavioral prerequisites, at least, for the shared attention level seem satisfied. A main challenge for the development of future social robots is to work on the ability to acquire common, mutual, and shared knowledge, to interpret others' ambiguous verbal or subtle nonverbal cues based on common ground in more complex environments, and to track shared experiences and integrate them into future interactions with others (for a review, see Yang et al., 2018). Another challenge is to identify what types of cues social robots and avatars need to produce in order to create at least an illusion of a second-person, engaged relation that humans will react to naturally (see, e.g., Anzalone, Boucenna, Ivaldi, & Chetouani, 2015; Breazeal & Scassellati, 2002; Castro-González, Admoni, & Scassellati, 2016; Lee, Breazeal, & DeSteno, 2017; Leite et al., 2017; Salem, Kopp, Wachsmuth, Rohlfing, & Joublin, 2012).

6.2. Application of this typology to joint goals and other joint mental states

We focus on attention and knowledge states here; however, the same typology (with the same levels of knowledge) can be extended to other mental states that can be either individual or joint (such as emotions, goals, desires, and beliefs) as well. Here we briefly sketch how the levels could apply to joint goals as an example (since similar definitional and terminological issues as in the field of joint attention apply to the field of joint action; Milward & Carpenter, 2018). To

achieve an *individual goal*, one intentionally organizes one's own behavior to reach that goal. For example, while reading the newspaper, Maria finds that she is strongly opposed to a law that has recently been passed. She sets for herself the goal of trying to convince the president to veto the law. She decides to stand in front of the president's residence and hold up a sign protesting against the controversial law.

One can also *monitor* another's goal, that is, infer someone else's goal, adopt it as one's own, and intentionally organize one's behavior to reach it. For example, Maria overhears a colleague mentioning holding up a sign in front of the president's residence. She adopts this goal, knowing that her colleague has the same goal.

In contrast, two or more individuals can have the same goal and know this together. Again, we propose that it is useful to distinguish several types of such goals based on the ways in which the goals are created: via inference in a third-person context or via direct experience with or without communication in a second-person context. Consider this example of a *common goal*: Maria stands in front of the president's residence with her sign. She sees that there are already several other people facing her holding similar signs on the other side of the residence, but she cannot go there as the police have blocked the streets. Maria realizes that they have the same goal, and that each of them will assume that they have a common goal with her. In a third-person situation, individuals can have a common goal by making inferences about the situation and/or engaging in recursive reasoning. This can enable them to coordinate their actions towards their common goal, even in situations in which no direct contact is possible.

To illustrate a *mutual goal*: Maria approaches the president's residence with her sign, and she makes eye contact with the protesters already standing there. They quickly gaze away as they need to be vigilant about the police's action. This mutual contact directly manifests to each of

them that they realize that they have a mutual goal to convince the president to veto the law. Two individuals can have a mutual goal when they become directly and reciprocally aware, via attention contact in a second-person relation but without intentional communication, of their same goal.

Once Maria talks to one of the protesters, “Let’s do this, veto the law!”, and he shouts back, “Yes!”, they establish a *shared goal*. Two individuals can establish a shared goal when they are in a second-person relation to each other, while they are engaged in bidirectional intentional communication about the goal.

In contrast to individual and monitored goals, common, mutual, and shared goals can give rise to additional, social obligations beyond one’s own individual commitment to achieve the goal. In common goals this additional commitment comes from the fact that each of the individuals knows about their knowledge of the goal and thus plausible deniability about the goal is lower. The personal connection in the mutual goal level makes the commitment to achieve the goal feel stronger than in the common goal level, and the public nature of the communication in the shared goal level goes beyond this to enable both individuals to acknowledge the shared goal and their shared knowledge of it, thus not only highlighting their individual commitment to achieve the goal, but also creating interpersonal, joint commitments to each other about the joint goal (Siposova et al., 2018). That is, as soon as it turns into any type of joint goal, by definition (Gilbert, 1990) there is some joint commitment to it – and the strength of the joint commitment lies on a continuum just like with social attention and social knowledge. Joint commitments support the achievement of joint goals, as they reduce uncertainty and we put more effort into fulfilling joint commitments (Michael & Pacherie, 2015).

Extending the typology to other mental states such as emotions, goals, desires, and beliefs

thus allows us to use consistent terminology across different mental states and start to bridge the different literatures on joint attention and joint action. It also accounts for both the observational (third person) and the experiential (second person) ways in which we understand and share the world with others.

6.3. Other possible levels

One final direction for future research and thinking is to consider whether, in addition to the levels described above, there might be further levels of social attention and knowledge, for example, a cultural or collective level in which we pay attention to the attention not only of our immediate social partners but also of the group as a whole.

7. Conclusion

We argue here that we will be able to address the question of what joint attention is and how the jointness in it is achieved more productively if we distinguish several levels of social attention. We offer a systematic framework of social attention and knowledge levels in which we distinguish between two types of relations: reasoning about others as an outside observer from a third-person perspective versus directly engaging with them within a second-person perspective. A key part of our proposal is the claim that we are not jointly attending to something with someone until we enter into a second-person relation with that person. This usually involves communication (including simple communicative looks) – our shared attention level – but it does not always: Neutral, non-communicative eye contact (even accidental eye contact, if it is registered) can create a bidirectional connection (in our mutual attention level). This proposal contributes to both the psychological and the philosophical literature by suggesting that social attention is not a single process, but rather it consists of a cluster of different cognitive skills. This proposal also emphasizes the link between the different attention levels and their

corresponding knowledge states and points to a need to study them together. We have pointed out that not only do the levels themselves lie on a scale of jointness, but there is a continuum of jointness within each of the social levels as well. This, as well as the fact that the levels can be nested hierarchically within a single episode, illustrates that social attention is far more complex than previously thought. These levels allow us to clearly see the gaps in our knowledge about social attention in different populations. It is also useful to distinguish these levels because of the different consequences they have in terms of what kinds of interactions and obligations they can support. First, monitoring attention itself is already an advanced social-cognitive skill, which gives advantages to the individuals capable of it, for example in making informed predictions about others' future actions. Second, the common, mutual, and shared levels can support obligations and even commitments, each level with varying degrees of strength. Additionally, each of these levels presents a distinct way to solve social coordination dilemmas in different types of circumstances.

Social attention is an important topic in many disciplines, ranging from philosophy, to developmental, comparative, and social psychology, to psychopathology, to social robotics. To be able to compare new findings and exchange knowledge across these diverse disciplines effectively, and bridge the literatures on joint attention and joint action, we need to be sure that we are talking about the same psychological processes. With this in mind, we propose that carefully distinguishing between social attention and knowledge levels helps us to study this intriguing topic better.

8. Acknowledgments

We thank Michael Tomasello for many years of extremely helpful discussions on this topic. We also thank Nathan Caruana, Jan Engelmann, Juan-Carlos Gómez, Sebastian Grueneisen, Margherita Isella, Lucy Liu, Cathal O'Madagain, Elisabeth Pacherie, Gideon Salter, Theo Toppe, Wouter Wolf, Dan Zahavi, and two anonymous reviewers for their useful feedback and comments.

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

9. References

Admoni, H., & Scassellati, B. (2017). Social eye gaze in human-robot interaction: A review. *Journal of Human-Robot Interaction*, *6*(1), 25–63. doi:10.5898/jhri.6.1.admoni

Akhtar, N., Jipson, J., & Callanan, M. A. (2001). Learning words through overhearing. *Child Development*, *72*(2), 416–430. doi:10.1111/1467-8624.00287

Amici, F., Aureli, F., Visalberghi, E., & Call, J. (2009). Spider monkeys (*Ateles geoffroyi*) and capuchin monkeys (*Cebus apella*) follow gaze around barriers: Evidence for perspective taking? *Journal of Comparative Psychology*, *123*(4), 368–374. doi:10.1037/a0017079

Anzalone, S. M., Boucenna, S., Ivaldi, S., & Chetouani, M. (2015). Evaluating the engagement with social robots. *International Journal of Social Robotics*, *7*(4), 465–478. doi:10.1007/s12369-015-0298-7

Anzalone, S. M., Tilmont, E., Boucenna, S., Xavier, J., Jouen, A.-L., Bodeau, N., … Cohen, D. (2014). How children with autism spectrum disorder behave and explore the 4-dimensional (spatial 3D+time) environment during a joint attention induction task with a robot. *Research in Autism Spectrum Disorders*, *8*(7), 814–826. doi:10.1016/j.rasd.2014.03.002

Argyle, M., & Cook, M. (1976). *Gaze and mutual gaze*. Cambridge, England: Cambridge University Press.

Bach, K., & Hamish, R. M. (1979). *Linguistics communication and speech acts*. Cambridge, MA: MIT Press.

Bakeman, R., & Adamson, L. B. (1984). Coordinating attention to people and objects in mother-infant and peer-infant interaction. *Child Development*, *55*(4), 1278–1289. doi:10.2307/1129997

Bard, K. A., Bakeman, R., Boysen, S. T., & Leavens, D. A. (2014). Emotional engagements predict and enhance social cognition in young chimpanzees. *Developmental Science*, *17*(5), 682–696. doi:10.1111/desc.12145

Bard, K. A., & Vauclair, J. (1984). The communicative context of object manipulation in ape and human adult-infant pairs. *Journal of Human Evolution*, *13*(2), 181–190. doi:10.1016/s0047-2484(84)80062-7

Baron-Cohen, S. (1989). Joint-attention deficits in autism: Towards a cognitive analysis. *Development and Psychopathology*, *1*(3), 185–189.

Bigelow, A. E. (2003). The development of joint attention in blind infants. *Development and Psychopathology*, *15*(2), 259–275. doi:10.1017/s0954579403000142

Bohn, M., Call, J., & Tomasello, M. (2016). The role of past interactions in great apes' communication about absent entities. *Journal of Comparative Psychology*, *130*(4), 351–357. doi:10.1037/com0000042

Boothby, E. J., Clark, M. S., & Bargh, J. A. (2014). Shared experiences are amplified. *Psychological Science*, *25*(12), 2209–2216. doi:10.1177/0956797614551162

Boothby, E. J., Smith, L. K., Clark, M. S., & Bargh, J. A. (2016). Psychological distance moderates the amplification of shared experience. *Personality and Social Psychology Bulletin*, *42*(10), 1431–1444. doi:10.1177/0146167216662869

Botero, M. (2016). Tactless scientists: Ignoring touch in the study of joint attention. *Philosophical Psychology*, *29*(8), 1200–1214. doi:10.1080/09515089.2016.1225293

Bräuer, J., Call, J., & Tomasello, M. (2005). All great ape species follow gaze to distant locations and around barriers. *Journal of Comparative Psychology*, *119*(2), 145–154. doi:10.1037/0735-7036.119.2.145

Bräuer, J., Call, J., & Tomasello, M. (2007). Chimpanzees really know what others can see in a competitive situation. *Animal Cognition*, *10*(4), 439–448. doi:10.1007/s10071-007-0088-1

Breazeal, C., & Scassellati, B. (2002). Robots that imitate humans. *Trends in Cognitive Sciences*, *6*(11), 481–487. doi:10.1016/s1364-6613(02)02016-8

Brinck, I. (2004). The pragmatics of imperative and declarative pointing. *Cognitive Science Quarterly*, *3*(4), 429–446.

Bruinsma, Y., Koegel, R. L., & Koegel, L. K. (2004). Joint attention and children with autism: A review of the literature. *Mental Retardation and Developmental Disabilities Research Reviews*, *10*(3), 169–175. doi:10.1002/mrdd.20036

Bruner, J. S. (1974). From communication to language - A psychological perspective. *Cognition*, *3*(3), 255–287. doi:10.1016/0010-0277(74)90012-2

Buber, M. (1958). *I and Thou*. New York, NY: Charles Scribner's Sons.

Bugnyar, T., Stowe, M., & Heinrich, B. (2004). Ravens, *Corvus corax*, follow gaze direction of humans around obstacles. *Proceedings of the Royal Society B: Biological Sciences*, *271*(1546), 1331–1336. doi:10.1098/rspb.2004.2738

Burghardt, G. M., & Greene, H. W. (1988). Predator simulation and duration of death feigning in neonate hognose snakes. *Animal Behaviour*, *36*(6), 1842–1844. doi:10.1016/s0003-3472(88)80127-1

Butterworth, G. (1995). Origins of minds in perception and action. In C. Moore, & P. J. Dunham (Eds.), *Joint attention: Its origins and role in development* (pp. 29–40). Hillsdale, NJ: Erlbaum.

Calabi, C. (2008). Winks, sighs and smiles? Joint attention, common knowledge and ephemeral groups. In N. Psarros, H. B. Schmidt, & K. Schulte-Ostermann (Eds.), *Concepts of sharedness* (p. 41–58). Frankfurt, Germany: Ontos Verlag.

Camaioni, L. (1993). The development of intentional communication: A re-analysis. In J. Nadel, & L. Camaioni (Eds.), *New perspectives in early communicative development* (pp. 82–96). London, England: Routledge.

Campbell, J. (2005). Joint attention and common knowledge. In N. Eilan, C. Hoerl, T. McCormack, & J. Roessler (Eds.), *Joint attention: Communication and other minds* (pp. 287–297). Oxford, England: Clarendon Press.

Carpenter, M., & Call, J. (2013). How joint is the joint attention of apes and human infants? In J. Metcalfe, & H.S. Terrace (Eds.), *Agency and joint attention* (pp. 49–61). New York, NY: Oxford University Press. doi:10.1093/acprof:oso/9780199988341.003.0003

Carpenter, M., & Liebal, K. (2011). Joint attention, communication, and knowing together in infancy. In A. Seemann (Ed.), *Joint attention: New developments in psychology, philosophy of mind, and social neuroscience* (pp. 159–181). Cambridge, MA: MIT Press.

Carpenter, M., Nagell, K., & Tomasello, M. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs of the Society for Research in Child Development, 63*(4), serial no. 255.

Carpenter, M., Tomasello, M., & Savage-Rumbaugh, S. (1995). Joint attention and imitative learning in children, chimpanzees, and enculturated chimpanzees. *Social Development*, *4*(3), 217–237. doi:10.1111/j.1467-9507.1995.tb00063.x

Castro-González, Á., Admoni, H., & Scassellati, B. (2016). Effects of form and motion on judgments of social robots' animacy, likability, trustworthiness and unpleasantness. *International Journal of Human-Computer Studies*, *90*, 27–38. doi:10.1016/j.ijhcs.2016.02.004

Chwe, M. S.-Y. (2013). *Rational ritual: Culture, coordination, and common knowledge*. Princeton, NJ: Princeton University Press.

Clark, H. H. (1992). *Arenas of language use*. Chicago, IL: University of Chicago Press.

Clark, H. H. (1996). *Using language*. Cambridge, England: Cambridge University Press.

Clark, H. H., & Marshall, C. R. (1981). *Definite reference and mutual knowledge*. Cambridge, England: Cambridge University Press.

Crockford, C., Wittig, R., Mundry, R. & Zuberbühler, K. (2012). Wild chimpanzees inform ignorant group members of danger. *Current Biology*, *22*(2), 142–146. doi:10.1016/j.cub.2011.11.053

Deák, G. O., Flom, R. A., & Pick, A. D. (2000). Effects of gesture and target on 12-and 18-month-olds' joint visual attention to objects in front of or behind them. *Developmental Psychology*, *36*(4), 511. doi:10.1037/0012-1649.36.4.511

DeJaegher, H., DiPaolo, E., & Gallagher, S. (2010). Can social interaction constitute social cognition? *Trends in Cognitive Sciences*, *14*(10), 441–447. doi:10.1016/j.tics.2010.06.009

Duguid, S., Wyman, E., Bullinger, A. F., Herfurth-Majstorovic, K., & Tomasello, M. (2014). Coordination strategies of chimpanzees and human children in a Stag Hunt game. *Proceedings of the Royal Society of London B: Biological Sciences*, *281*(1796). doi:10.1098/rspb.2014.1973

Eilan, N. (submitted). Joint attention and the second person. *Manuscript submitted for publication. Draft available online: http://www2.warwick.ac.uk/fac/soc/philosophy/people/eilan/*.

Emery, N. J. (2000). The eyes have it: the neuroethology, function and evolution of social gaze. *Neuroscience and Biobehavioral Reviews*, *24*(6), 581–604. doi:10.1016/s0149-7634(00)00025-7

Farroni, T., Csibra, G., Simion, F., & Johnson, M. H. (2002). Eye contact detection in humans from birth. *Proceedings of the National Academy of Sciences*, *99*(14), 9602–9605. doi:10.1073/pnas.152159999

Farroni, T., Mansfield, E. M., Lai, C., & Johnson, M. H. (2003). Infants perceiving and acting on the eyes: Tests of an evolutionary hypothesis. *Journal of Experimental Child Psychology*, *85*(3), 199–212. doi:10.1016/s0022-0965(03)00022-5

Floor, P., & Akhtar, N. (2006). Can 18-month-old infants learn words by listening in on conversations? *Infancy*, *9*(3), 327–339. doi:10.1207/s15327078in0903_4

Flombaum, J. I., & Santos, L. R. (2005). Rhesus monkeys attribute perceptions to others. *Current Biology*, *15*(5), 447–452. doi:10.1016/j.cub.2004.12.076

Fuchs, T. (2012). The phenomenology and development of social perspectives. *Phenomenology and the Cognitive Sciences*, *12*(4), 655–683. doi:10.1007/s11097-012-9267-x

Gallotti, M., & Frith, C. D. (2013). Social cognition in the we-mode. *Trends in Cognitive Sciences*, *17*(4), 160–165. doi:10.1016/j.tics.2013.02.002

Ganea, P. A., & Saylor, M. M. (2007). Infants' use of shared linguistic information to clarify ambiguous requests. *Child Development*, *78*(2), 493–502. doi:10.1111/j.1467-8624.2007.01011.x

Genty, E., Neumann, C., & Zuberbühler, K. (2015). Bonobos modify communication signals according to recipient familiarity. *Scientific Reports*, *5*(1). doi:10.1038/srep16442

Geurts, B. (in press). Communication as commitment making: speech acts, implicatures, common ground. *Theoretical Linguistics*.

Gilbert, M. (1990). Walking together: A paradigmatic social phenomenon. *Midwest Studies In Philosophy*, *15*(1), 1–14. doi:10.1111/j.1475-4975.1990.tb00202.x

Gilbert, M. (2007). Mutual recognition, common knowledge, and joint attention. In T. Ronnow-Rasmussen, B. Petersson, J. Josefsson, & D. Egonsson (Eds.), *Hommage à Wlodek: Philosophical papers dedicated to Wlodek Rabinowicz* (pp. 1–21). Lund, Sweden: Lund University.

Gómez, J. C. (1996). Ostensive behavior in great apes: The role of eye contact. In A. E. Russon, K. A. Bard, & S. T. Parker (Eds.), *Reaching into thought: The minds of the great apes* (pp. 131–151). Cambridge, England: Cambridge University Press.

Gómez, J. C. (2005). Joint attention and the notion of subject: Insights from apes, normal children and children with autism. In N. Eilan, C. Hoerl, T. McCormack, & J. Roessler (Eds.), *Joint attention: Communication and other minds* (pp. 65–84). Oxford, England: Clarendon Press.

Gómez, J. C., Sarriá, E., & Tamarit, J. (1993). The comparative study of early communication and theories of mind: Ontogeny, phylogeny, and pathology. In S. Baron-Cohen, H. Tager-Flusberg, & D. J. Cohen (Eds.), *Understanding other minds: Perspectives from autism* (pp. 397–426). New York, NY: Oxford University Press.

Grosse, G., Scott-Phillips, T. C., & Tomasello, M. (2013). Three-year-olds hide their communicative intentions in appropriate contexts. *Developmental Psychology*, *49*(11), 2095–2101. doi:10.1037/a0032017

Grueneisen, S., Wyman, E., & Tomasello, M. (2015). Children use salience to solve coordination problems. *Developmental Science*, *18*(3), 495–501. doi:10.1111/desc.12224

Hains, S. M. J., & Muir, D. W. (1996). Infant sensitivity to adult eye direction. *Child Development*, *67*(5), 1940–1951. doi:10.2307/1131602

Hare, B., Call, J., Agnetta, B., & Tomasello, M. (2000). Chimpanzees know what conspecifics do and do not see. *Animal Behaviour*, *59*(4), 771–785. doi:10.1006/anbe.1999.1377

Harder, P., & Kock, C. (1976). *The theory of presupposition failure*. Copenhagen, Denmark: Akademisk Forlag.

Haun, D. B. M., & Over, H. (2013). Like me: A homophily-based account of human culture. In P. J. Richerson, & M. H. Christiansen (Eds.), *Cultural evolution: Society, technology, language, and religion* (pp. 75–85). Cambridge, MA: MIT.

Heal, J. (2005). Joint attention and understanding the mind. In N. Eilan, C. Hoerl, T. McCormack, & J. Roessler (Eds.), *Joint attention: Communication and other minds*. Oxford, England: Clarendon Press.

Hennig, C. W. (1977). Effects of simulated predation on tonic immobility in *Anolis carolinensis*: The role of eye contact. *Bulletin of the Psychonomic Society*, *9*(4), 239–242. doi:10.3758/bf03336987

Hobson, J. A., & Hobson, R. P. (2007). Identification: The missing link between joint attention and imitation? *Development and Psychopathology*, *19*(02), 411–431. doi:10.1017/s0954579407070204

Hobson, R. P. (1993). The emotional origins of social understanding. *Philosophical Psychology*, *6*(3), 227–249. doi:10.1080/09515089308573090

Hobson, R. P. (2005). What puts the jointness into joint attention? In N. Eilan, C. Hoerl, T. McCormack, & J. Roessler (Eds.), *Joint attention: Communication and other minds* (pp. 185–204). Oxford, England: Clarendon Press.

Hobson, R. P. (2007). Communicative depth: Soundings from developmental psychopathology. *Infant Behavior and Development*, *30*(2), 267–277. doi:10.1016/j.infbeh.2007.02.006

Hobson, R. P. (2014). The coherence of autism. *Autism*, *18*(1), 6–16. doi:10.1177/1362361313497538

Hobson, R. P., & Hobson, J. A. (2008). Engaging, sharing, knowing. Some lessons from research in autism. In J. Zlatev, T. P. Racine, C. Sinha, & E. Itkonen (Eds.), *The shared mind. Perspectives on intersubjectivity (pp. 67–88).* Amsterdam, Netherlands: John Benjamins.

Hood, B. M., Willen, J. D., & Driver, J. (1998). Adult's eyes trigger shifts of visual attention in human infants. *Psychological Science*, *9*(2), 131–134. doi:10.1111/1467-9280.00024

Kaminski, J., Call, J., & Tomasello, M. (2008). Chimpanzees know what others know, but not what they believe. *Cognition*, *109*(2), 224–234. doi:10.1016/j.cognition.2008.08.010

Kaminski, J., Pitsch, A., & Tomasello, M. (2013). Dogs steal in the dark. *Animal Cognition*, *16*(3), 385–394. doi:10.1007/s10071-012-0579-6

Kaminski, J., Riedel, J., Call, J., & Tomasello, M. (2005). Domestic goats, *Capra hircus*, follow gaze direction and use social cues in an object choice task. *Animal Behaviour*, *69*(1), 11–18. doi:10.1016/j.anbehav.2004.05.008

Kaplan, F., & Hafner, V. V. (2006). The challenges of joint attention. *Interaction Studies*, *7*(2), 135–169.

Leavens, D. A., & Bard, K. A. (2011). Environmental influences on joint attention in great apes: implications for human cognition. *Journal of Cognitive Education and Psychology*, *10*(1), 9–31. doi:10.1891/1945-8959.10.1.9

Leavens, D. A., Hopkins, W. D., & Bard, K. A. (2005). Understanding the point of chimpanzee pointing. *Current Directions in Psychological Science*, *14*(4), 185–189. doi:10.1111/j.0963-7214.2005.00361.x

Leavens, D. A., & Racine, T. (2009). Joint attention in apes and humans. Are humans unique? *Journal of Consciousness Studies*, *16*, 240–267.

Lee, J. J., Breazeal, C., & DeSteno, D. (2017). Role of speaker cues in attention inference. *Frontiers in Robotics and AI*, *4*. doi:10.3389/frobt.2017.00047

Leite, I., McCoy, M., Lohani, M., Ullman, D., Salomons, N., Stokes, C., … Scassellati, B. (2017). Narratives with robots: The impact of interaction context and individual differences on story recall and emotional understanding. *Frontiers in Robotics and AI*, *4*. doi:10.3389/frobt.2017.00029

Leekam, S., Baron-Cohen, S., Perrett, D., Milders, M., & Brown, S. (1997). Eye-direction detection: A dissociation between geometric and joint attention skills in autism. *British Journal of Developmental Psychology*, *15*(1), 77–95. doi:10.1111/j.2044-835x.1997.tb00726.x

Lemaignan, S., Garcia, F., Jacq, A., & Dillenbourg, P. (2016). From real-time attention assessment to with-me-ness in human-robot interaction. In *The Eleventh ACM/IEEE International Conference on Human Robot Interaction* (pp. 157–164). IEEE Press. doi:10.1109/hri.2016.7451747

Lewis, D. (1969). *Convention: A philosophical study*. Cambridge, MA: Harvard University Press.

Lewis, D. (1979). Scorekeeping in a language game. *Journal of Philosophical Logic*, *8*(1), 339–359. doi:10.1007/bf00258436

Liebal, K., Behne, T., Carpenter, M., & Tomasello, M. (2009). Infants use shared experience to interpret pointing gestures. *Developmental Science*, *12*(2), 264–271. doi:10.1111/j.1467-7687.2008.00758.x

Liebal, K., Carpenter, M., & Tomasello, M. (2010). Infants' use of shared experience in declarative pointing. *Infancy*, *15*(5), 545–556. doi:10.1111/j.1532-7078.2009.00028.x

Liszkowski, U., Carpenter, M., Henning, A., Striano, T. & Tomasello, M. (2004). Twelve-month-olds point to share attention and interest. *Developmental Science*, *7*(3), 297–307. doi:10.1111/j.1467-7687.2004.00349.x

Melis, A. P., Call, J., & Tomasello, M. (2006). Chimpanzees (*Pan troglodytes*) conceal visual and auditory information from others. *Journal of Comparative Psychology*, *120*(2), 154–162. doi:10.1037/0735-7036.120.2.154

Met, A., Miklósi, Á., & Lakatos, G. (2014). Gaze-following behind barriers in domestic dogs. *Animal Cognition*, *17*(6), 1401–1405. doi:10.1007/s10071-014-0754-z

Michael, J. (2011). Interactionism and mindreading. *Review of Philosophy and Psychology*, *2*(3), 559–578. doi:10.1007/s13164-011-0066-z

Michael, J., & Pacherie, E. (2015). On commitments and other uncertainty reduction tools in joint action. *Journal of Social Ontology*, *1*(1), 89–120.

Miklósi, Á., Polgárdi, R., Topál, J., & Csányi, V. (1998). Use of experimenter-given cues in dogs. *Animal Cognition*, *1*(2), 113–121. doi:10.1007/s100710050016

Milward, S., & Carpenter, M. (2018). Joint action and joint attention: drawing parallels between the literatures. *Social and Personality Psychology Compass, 12(4),* e12377.

Moll, H., Carpenter, M., & Tomasello, M. (2007). Fourteen-month-olds know what others experience only in joint engagement. *Developmental Science*, *10*(6), 826–835. doi:10.1111/j.1467-7687.2007.00615.x

Moll, H., Richter, N., Carpenter, M., & Tomasello, M. (2008). Fourteen-month-olds know what "we" have shared in a special way. *Infancy*, *13*(1), 90–101. doi:10.1080/15250000701779402

Moll, H., & Tomasello, M. (2004). 12- and 18-month-old infants follow gaze to spaces behind barriers. *Developmental Science*, *7*(1), 1–9. doi:10.1111/j.1467-7687.2004.00315.x

Moll, H. & Tomasello, M. (2007a). Cooperation and human cognition: the Vygotskian intelligence hypothesis. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *362*(1480), 639–648. doi:10.1098/rstb.2006.2000

Moll, H. & Tomasello, M. (2007b). How 14- and 18-month-olds know what others have experienced. *Developmental Psychology*, *43*(2), 309–317. doi:10.1037/0012-1649.43.2.309

Monderer, D., & Samet, D. (1989). Approximating common knowledge with common beliefs. *Games and Economic Behavior*, *1*(2), 170–190.

Mundy, P. (2003). Annotation: The neural basis of social impairments in autism: the role of the dorsal medial-frontal cortex and anterior cingulate system. *Journal of Child Psychology and Psychiatry*, *44*(6), 793–809. doi:10.1111/1469-7610.00165

Mundy, P., & Newell, L. (2007). Attention, joint attention, and social Cognition. *Current Directions in Psychological Science*, *16*(5), 269–274. doi:10.1111/j.1467-8721.2007.00518.x

Mundy, P., Sigman, M., & Kasari, C. (1990). A longitudinal study of joint attention and language development in autistic children. *Journal of Autism and Developmental Disorders*, *20*(1), 115–128. doi:10.1007/bf02206861

Mundy, P., Sullivan, L., & Mastergeorge, A. M. (2009). A parallel and distributed-processing model of joint attention, social cognition and autism. *Autism Research*, *2*(1), 2–21. doi:10.1002/aur.61

Mutlu, B., Yamaoka, F., Kanda, T., Ishiguro, H., & Hagita, N. (2009). Nonverbal leakage in robots: Communication of intentions through seemingly unintentional behavior. In *Proceedings of the 4th ACM/IEEE international conference on Human robot interaction* (pp. 69–76). ACM Press. doi:10.1145/1514095.1514110

O'Madagain, C., & Tomasello, M., in press. Joint attention to mental content and the social origin of reasoning. *Synthese*.

Pack, A. A., & Herman, L. M. (2006). Dolphin social cognition and joint attention: Our current understanding. *Aquatic Mammals*, *32*(4), 443–460. doi:10.1578/am.32.4.2006.443

Peacocke, C. (2005). Joint attention: Its nature, reflexivity, and relation to common knowledge. In N. Eilan, C. Hoerl, T. McCormack, & J. Roessler (Eds.), *Joint attention: Communication and other minds* (pp. 298–324). Oxford, England: Clarendon Press.

Pinker, S. (2007). *The stuff of thought: Language as a window into human nature*. New York, NY: Viking.

Phillips, W., Gómez, J. C., Baron-Cohen, S., Laá, V., & Rivière, A. (1995). Treating people as objects, agents, or "subjects": How young children with and without autism make requests. *Journal of Child Psychology and Psychiatry*, *36*(8), 1383–1398. doi:10.1111/j.1469-7610.1995.tb01670.x

Proops, L., & McComb, K. (2009). Attributing attention: the use of human-given cues by domestic horses (*Equus caballus*). *Animal Cognition*, *13*(2), 197–205. doi:10.1007/s10071-009-0257-5

Range, F., & Virányi, Z. (2011). Development of gaze following abilities in wolves (*Canis Lupus*). *PLoS ONE*, *6*(2), 1–9. doi:10.1371/journal.pone.0016888

Reddy, V. (1996). Omitting the second person in social understanding. *Behavioral and Brain Sciences*, *19*(01), 140. doi:10.1017/s0140525x00041996

Reddy, V. (2003). On being the object of attention: Implications for self-other consciousness. *Trends in Cognitive Sciences*, *7*(9), 397–402. doi:10.1016/s1364-6613(03)00191-8

Reddy, V. (2011). A gaze at grips with me. In A. Seemann (Ed.), *Joint attention: New developments in psychology, philosophy of mind, and social neuroscience* (pp. 137–159). Cambridge, MA: MIT Press.

Reddy, V. (2018). Why engagement? A second person take on social cognition. In A. Newen, L. de Bruin, & S. Gallagher (Eds.), *Oxford handbook of 4e cognition*. Oxford, England: Oxford University Press.

Reddy, V., & Morris, P. (2004). Participants don't need theories: Knowing minds in engagement. *Theory, & Psychology*, *14*(5), 647 – 665. doi:10.1177/0959354304046177

Ristau, C. A. (1991). Before mindreading: Attention, purposes and deception in birds? In A. Whiten (Ed.), *Natural theories of mind. Evolution, development and simulation of mindreading* (pp. 209–222). Cambridge, MA: Basil Blackwell.

Rivas, E. (2005). Recent use of signs by chimpanzees (*Pan troglodytes*) in interactions with humans. *Journal of Comparative Psychology*, *119*(4), 404–417. doi:10.1037/0735-7036.119.4.404

Rossmanith, N., Costall, A., Reichelt, A. F., López, B., & Reddy, V. (2014). Jointly structuring triadic spaces of meaning and action: book sharing from 3 months on. *Frontiers in Psychology*, *5*, 1–22. doi:10.3389/fpsyg.2014.01390

Salem, M., Kopp, S., Wachsmuth, I., Rohlfing, K., & Joublin, F. (2012). Generation and evaluation of communicative robot gesture. *International Journal of Social Robotics*, *4*(2), 201–217. doi:10.1007/s12369-011-0124-9

Santos, L. R., Nissen, A. G., & Ferrugia, J. A. (2006). Rhesus monkeys, *Macaca mulatta*, know what others can and cannot hear. *Animal Behaviour*, *71*(5), 1175–1181. doi:10.1016/j.anbehav.2005.10.007

Saylor, M. M., & Ganea, P. (2007). Infants interpret ambiguous requests for absent objects. *Developmental Psychology*, *43*(3), 696–704. doi:10.1037/0012-1649.43.3.696

Senju, A., & Csibra, G. (2008). Gaze following in human infants depends on communicative signals. *Current Biology*, *18*(9), 668–671. doi:10.1016/j.cub.2008.03.059

Scaife, M., & Bruner, J. S. (1975). The capacity for joint visual attention in the infant. *Nature*, *253*(5489), 265–266. doi:10.1038/253265a0

Schachter, S. (1959). *The psychology of affiliation*. Stanford, CA: Stanford University Press.

Schiffer, S. R. (1972). *Meaning*. Oxford, England: Clarendon Press.

Schilbach, L. (2010). A second-person approach to other minds. *Nature Reviews Neuroscience*, *11*(6), 449. doi:10.1038/nrn2805-c1

Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., & Vogeley, K. (2013). Toward a second-person neuroscience. *Behavioral and Brain Sciences*, *36*(4), 393–414.

Siposova, B., Tomasello, M., & Carpenter, M. (2018). Communicative eye contact signals a commitment to cooperate for young children. *Cognition*.

Sperber, D., & Wilson, D. (1986). *Relevance: Communication and cognition*. Oxford, England: Basil Blackwell.

Striano, T., & Bertin, E. (2005). Coordinated affect with mothers and strangers: A longitudinal analysis of joint engagement between 5 and 9 months of age. *Cognition and Emotion*, *19*(5), 781–790. doi:10.1080/02699930541000002

Tanner, J. E., & Byrne, R. W. (2010). Triadic and collaborative play by gorillas in social games with objects. *Animal Cognition*, *13*(4), 591–607. doi:10.1007/s10071-009-0308-y

Taylor, C. (1985). *Human agency and language*. Cambridge, England: Cambridge University Press.

Thomas, K. A., DeScioli, P., Haque, O. S., & Pinker, S. (2014). The psychology of coordination and common knowledge. *Journal of Personality and Social Psychology*, *107*(4), 657–676. doi:10.1037/a0037037

Tomasello, M. (1995). Joint attention as social cognition. In C. Moore, & P. Dunham (Eds.), *Joint attention: Its origins and role in development* (pp. 103–130). Hillsdale, NJ: Erlbaum.

Tomasello, M. (1999). *The cultural origins of human cognition*. Cambridge, MA: Harvard University Press.

Tomasello, M. (2006). Why don't apes point? In N. Enfield, & S. Levinson (Eds.), *Roots of human sociality* (pp. 506–524). New York, NY: Wenner-Grenn.

Tomasello, M. (2008). *Origins of human communication*. Cambridge, MA: MIT Press.

Tomasello, M. (2011). Human culture in evolutionary perspective. In M. J. Gelfand, C. Y. Chiu, & Y. Y. Hong (Eds.), *Advances in culture and psychology: Vol. 1.* (pp. 5–51). New York, NY: Oxford University Press.

Tomasello, M., & Carpenter, M. (2005). The emergence of social cognition. *Monographs of the Society for Research in Child Development*, *70*, 1–122.

Tomasello, M. & Haberl, K. (2003). Understanding attention: 12- and 18-month-olds know what is new for other persons. *Developmental Psychology*, *39*(5), 906–912. doi:10.1037/0012-1649.39.5.906

Trevarthen, C. (1979). Communication and cooperation in early infancy: A description of primary intersubjectivity. In M. Bullowa (Ed.), *Before speech: The beginning of interpersonal communication* (pp. 321–347). New York, NY: Cambridge University Press.

Tomonaga, M., Tanaka, M., Matsuzawa, T., Myowa-Yamakoshi, M., Kosugi, D., Mizuno, Y., … Bard, K. A. (2004). Development of social cognition in infant chimpanzees (*Pan troglodytes*): Face recognition, smiling, gaze, and the lack of triadic interactions. *Japanese Psychological Research*, 46(3), 227–235. doi:10.1111/j.1468-5584.2004.00254.x

Vanderschraaf, P. & Sillari, G. (2014). Common knowledge. In E. Zalta (Ed.), *Stanford Encyclopedia of Philosophy*. Retrieved from https://plato.stanford.edu/archives/spr2014/entries/common-knowledge

Virányi, Z., Topál, J., Gácsi, M., Miklósi, Á., & Csányi, V. (2004). Dogs respond appropriately to cues of humans' attentional focus. *Behavioural Processes*, *66*(2), 161–172. doi:10.1016/j.beproc.2004.01.012

Vygotsky, L. (1978). Interaction between learning and development. *Readings on the Development of Children*, *23*(3), 34–41.

Warreyn, P., Roeyers, H., Wetswinkel, U. V., & Groote, I. D. (2006). Temporal coordination of joint attention behavior in preschoolers with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, *37*(3), 501–512. doi:10.1007/s10803-006-0184-0

Wilkinson, A., Mandl, I., Bugnyar, T., & Huber, L. (2010). Gaze following in the red-footed tortoise (*Geochelone carbonaria*). *Animal Cognition*, *13*(5), 765–769. doi:10.1007/s10071-010-0320-2

Wolf, W., Launay, J., & Dunbar, R. I. M. (2015). Joint attention, shared goals, and social bonding. *British Journal of Psychology*, *107*(2), 322–337. doi:10.1111/bjop.12144

Yang, G.-Z., Bellingham, J., Dupont, P. E., Fischer, P., Floridi, L., Full, R., … Wood, R. (2018). The grand challenges of science robotics. *Science Robotics*, *3*(14). doi:10.1126/scirobotics.aar7650

Zahavi, D. (2008). Simulation, projection and empathy. *Consciousness and Cognition*, *17*(2), 514–522. doi:10.1016/j.concog.2008.03.010

Zahavi, D. (2011). Empathy and direct social perception: A phenomenological proposal. *Review of Philosophy and Psychology*, *2*(3), 541–558. doi:10.1007/s13164-011-0070-3

Zahavi, D. (2015). Self and other: From pure ego to co-constituted we. *Continental Philosophy Review*, *48*(2), 143–160. doi:10.1007/s11007-015-9328-2

10. Conflicts of interest:

None.